

Review

N-Terminomics Strategies for Protease Substrates Profiling

Mubashir Mintoo , Amritangshu Chakravarty and Ronak Tilwala *

Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, USA; mubashirjaved63@ku.edu (M.M.); amritangshu.chakravarty@ku.edu (A.C.)

* Correspondence: ronak.tilwala@ku.edu; Tel.: +1-785-864-1020

Abstract: Proteases play a central role in various biochemical pathways catalyzing and regulating key biological events. Proteases catalyze an irreversible post-translational modification called proteolysis by hydrolyzing peptide bonds in proteins. Given the destructive potential of proteolysis, protease activity is tightly regulated. Dysregulation of protease activity has been reported in numerous disease conditions, including cancers, neurodegenerative diseases, inflammatory conditions, cardiovascular diseases, and viral infections. The proteolytic profile of a cell, tissue, or organ is governed by protease activation, activity, and substrate specificity. Thus, identifying protease substrates and proteolytic events under physiological conditions can provide crucial information about how the change in protease regulation can alter the cellular proteolytic landscape. In recent years, mass spectrometry-based techniques called N-terminomics have become instrumental in identifying protease substrates from complex biological mixtures. N-terminomics employs the labeling and enrichment of native and neo-N-termini peptides, generated upon proteolysis followed by mass spectrometry analysis allowing protease substrate profiling directly from biological samples. In this review, we provide a brief overview of N-terminomics techniques, focusing on their strengths, weaknesses, limitations, and providing specific examples where they were successfully employed to identify protease substrates in vivo and under physiological conditions. In addition, we explore the current trends in the protease field and the potential for future developments.

Keywords: protease substrates; N-terminomics; COFRADIC; TAILS; subtiligase; CHOPS



Citation: Mintoo, M.; Chakravarty, A.; Tilwala, R. N-Terminomics Strategies for Protease Substrates Profiling. *Molecules* **2021**, *26*, 4699. <https://doi.org/10.3390/molecules26154699>

Academic Editors: Sanjai Kumar Pathak and Dibyendu Dana

Received: 16 June 2021
Accepted: 22 July 2021
Published: 3 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proteases catalyze an irreversible post-translational modification called proteolysis in proteins by hydrolyzing peptide bonds [1]. More than 640 protease genes are discovered to date in the human genome, and they occupy about 3% of the entire genome [2,3]. Based on their catalytic activity, proteases are classified into six distinct classes: serine, cysteine, threonine, glutamic, aspartic, and metalloproteases [4]. Proteases are central to numerous biological pathways and play a vital role in regulating many signaling processes, such as cell-cycle progression [5], cell proliferation [6], cell death [7], DNA replication [8], tissue remodeling [9], hemostasis [10], wound healing [11], and immune responses [11]. Consistent with their essential roles, alterations in proteolytic systems result in multiple pathological conditions, including cancer [12], neurodegenerative disorders [13], infections [13], allergies [14], blood clotting disorders [15], and cardiovascular diseases [16]. Given the enormity of proteolytic potential, protease activity must be tightly regulated to prevent inappropriate and frequently destructive proteolysis. There are several mechanisms by which proteolytic activity is regulated in cells, such as regulation of gene expression [17,18], activation of their inactive zymogens [19–21], blockade by endogenous inhibitors [22,23], and via post-translational modifications such as glycosylation, proteolysis, and degradation [4].

Proteolytic cleavage results in the formation of two novel proteins/peptides, creating a new N- or C-terminus [24]. As a result, proteolysis changes the size, structure, and function of protein substrates. The proteolytic profile of a cell or organ is governed by protease activation, activity, and substrate specificity. Most proteases are secreted and

circulated as inactive zymogens and become activated by specific cleavage upon triggering events [25]. After activation, the activity of proteases is regulated by their endogenous protein inhibitors, post-translational modifications, or pH. For example, the activity of cathepsins in the lysosomes and pepsin in gastric secretions is pH dependent [26,27]. Additionally, protease expression can be transcriptionally regulated and restricted to a specific cell type [28]. Finally, protease substrate specificity is determined based on protease-substrate interactions. Proteases cleave their substrates by recognizing either single amino acid or a particular sequence of amino acids [29,30]. Proteases with narrow specificity generally execute limited proteolysis (e.g., caspases during apoptosis), while proteases with broad specificity (e.g., cysteine cathepsins) often have significant roles in general protein degradation and clearance. Optimal protease activation, activity, and substrate specificity determine the proteome-wide signaling network of the cells, tissues, and organs—any changes or abnormalities in proteolysis result in catastrophic outcomes.

Identifying protease substrates and proteolytic events under physiological conditions can provide crucial information about how the change in protease regulation alters the entire cellular proteolytic network. However, the identification of physiological protease substrates remained challenging for a long time due to their low abundance and the unstable nature of the newly generated peptide fragments, which can be further targeted for the subsequent proteolytic degradation [31]. Additionally, the cellular heterogeneity and changes in protease expression make it challenging to dissect the relevant proteolysis from background events in complex biological mixtures [32].

In recent years, mass spectrometry-based techniques have become instrumental in identifying protease substrates under physiological conditions [32,33]. These proteomic techniques are developed by targeting the newly formed N- or C-terminal of the protease substrates and are known as N- or C-terminomics, respectively [34,35]. However, since the carboxyl group is less reactive than primary amines, the amine labeling techniques, N-terminomics, are more successfully implemented and widely used. Herein, we provide a short overview of N-terminomics techniques developed in the past decade, focusing on their strengths, weaknesses, limitations, and providing specific examples where they were successfully employed to identify protease substrates *in vivo* and under physiological conditions. We further explore the current trends in the protease field and the potential for future developments.

2. Identification of Protease Substrates Using N-Terminomics Methods

N-terminomics methods detect the protease substrates from the complex mixture by identifying the native or neo-N-terminal generated upon proteolysis using LC-MS/MS analysis. In this review, we discuss the four major N-terminomics methods: Combinatorial Fractional Diagonal Chromatography (COFRADIC), Terminal Amine Isotopic Labeling of Substrates (TAILS), Subtiligase and Chemical enrichment of Protein Substrates (CHOPS). The most challenging step in the N-terminomics method is separating and isolating native or neo-N-terminal peptides from the internal tryptic peptides generated during the sample preparation (Figure 1). The tryptic peptides comprise more than 90% of total peptides and can easily outweigh the native or neo-N-terminal peptides in the sample. Based on how neo-N-terminal peptides are separated from the tryptic peptides, the N-terminomics methods can be further classified into negative enrichment methods and positive enrichment methods. In the negative enrichment methods (e.g., COFRADIC and TAILS), the tryptic peptides are enriched to isolate the native or neo-N-terminal peptides from the sample. While in the positive enrichment methods (e.g., Subtiligase, and CHOPS), native or neo-N-terminal peptides are directly enriched and isolated from the sample.

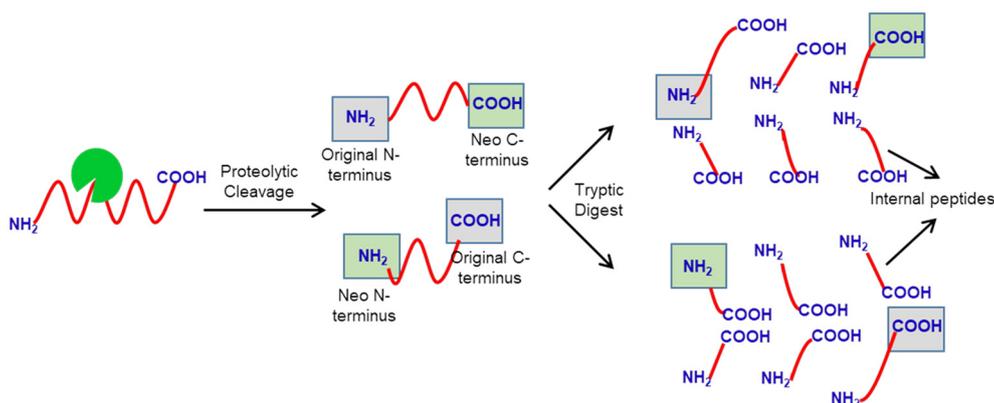


Figure 1. N-terminomics approach. Neo-N-terminal peptides and internal tryptic peptides. The overall workflow of the N-terminomics includes the separation of neo-N-terminal peptides from the terminal tryptic peptides.

N-terminomics methods can be further divided into *forward* N-terminomics and *reverse* N-terminomics [36]. The *forward* N-terminomics is used to identify the global proteolytic events, while the *reverse* N-terminomics is used to identify the proteolytic events occurring from a protease of interest (Figure 2). In the *forward* N-terminomics method, control and diseased/treated samples are lysed and subjected to N-terminomics workflow followed by LC-MS/MS analysis (Figure 2A). The identified peptides between two samples are then compared to dissect the differential proteolytic events associated with the treated/diseased samples. For example, COFRADIC, TAILS, and Subtiligase have been used to study Matrix Metallo Proteases (MMP) activity in cancer cells [37,38], cathepsin function of protein degradation, and proteolysis in pancreatic tumors and human blood samples [39]. The detailed list is compiled in Table 1. While the *forward* N-terminomics provides information about the relative proteolysis between two samples, it cannot identify specific protease substrates associated with the protease of interest. For this purpose, *reverse* N-terminomics methods are used. In this method, samples are lysed, the endogenous proteases are quenched, and native and neo-N-terminal peptides are blocked. Next, the samples are treated with the protease of interest. The control and treated samples are then subjected to N-terminomics workflow followed by LC-MS/MS analysis. Finally, the identified peptides from both samples are compared to dissect the physiological substrates of protease of interest (Figure 2B). The *reverse* N-terminomics has been successfully employed to identify physiological substrates of various proteases, including ADAMTS7 [40] and MMPs [41], using TAILS, granzyme tryptases, and several caspases using COFRADIC and Subtiligase [42,43]. The detailed list is compiled in Table 1. The *reverse* N-terminomics approach identifies physiological substrates of the target protease directly from the biological sample. Importantly, it also reveals the protease primary sequence specificity. However, *reverse* N-terminomics does not provide the functional or biological significance of the proteolytic events identified and can lead to the identification of substrates from organelles that would not normally co-localize with the protease in a living cell.

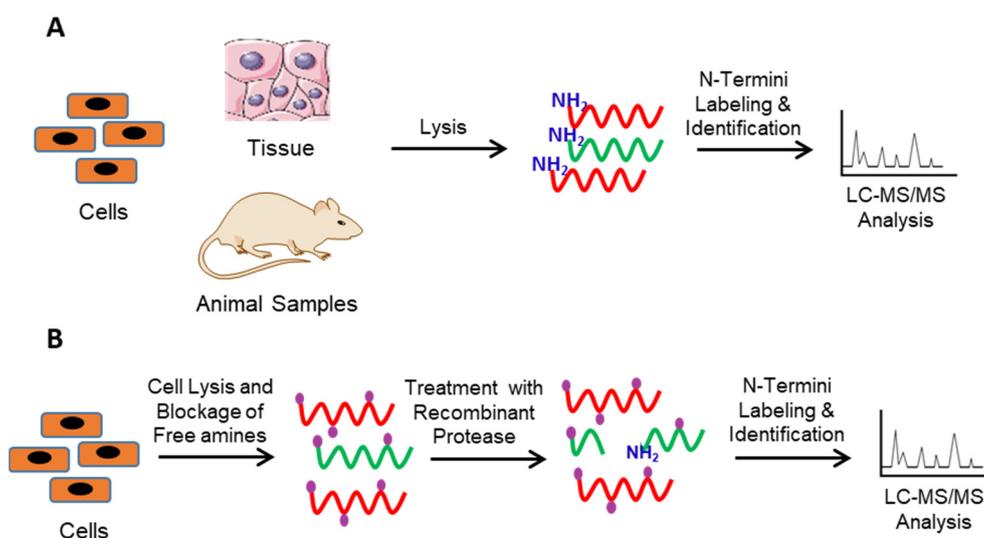


Figure 2. Forward and Reverse approaches of N-terminomics. (A) In the forward N-terminomics approach, treated and control samples lysed and subjected N-terminomics workflow followed by LC-MS/MS analysis to identify the global proteolysis. (B) In the reverse N-terminomics, samples are lysed, followed by the addition of protease of interest. The treated and control samples are then subjected to LC-MS/MS analysis to identify the physiological substrates and substrate specificity of the protease of interest.

Table 1. Applications of N-terminomics methods.

Method	Applications
COFRADIC	<ul style="list-style-type: none"> • Identification of the N-terminal acetylation sites in human and yeast proteins [44] • Identifying caspase cleavage sites during apoptosis [45] • Sequence specificity of initiator and executioner caspases [46] • Substrate identification for HIV-1 protease [47]
Subtiligase	<ul style="list-style-type: none"> • Identification of caspase substrates [48–54] • Identification of the Zika virus NS2B-NS3 protease substrates [55] • Protease cleavage sites on membranes and the surface of living cells [56] • Identification of unique proteins and peptides in the human serum and plasma [57] and discovery of disease-specific biomarkers [57–63]
TAILS	<ul style="list-style-type: none"> • Identification cathepsin D substrate profile from triple-negative breast cancer background [64] • Identification of proteolytic events from pancreatic tumors [39] • Identification of proteolytic events from ulcerative colitis [65] • Identification of proteolytic events from COPD [66] • Identification of MALT1 substrates in NF-κB mediated signaling in the presence of antigens [67] • Substrate profiling of DPP8 and DPP9 proteases [68] • Substrate profiling of metalloproteases [41,69–73] • Identification of proteolytic events and protease cleavage sites during Enterococcal <i>E. coli</i> [74] and viral infections [75]
CHOPS	<ul style="list-style-type: none"> • Substrate profiling of DPP8 and DPP9 proteases [76]

COFRADIC Combinatorial Fractional Diagonal Chromatography (COFRADIC) is one of the earlier N-terminomics methods [77]. This negative N-terminomics method uses different N-termini labeling to induce chromatographic shifts to isolate the tryptic peptides from the native or neo-N-terminal peptides. In this method, the native or neo-N-terminal peptides are blocked by acetylation followed by trypsin digestion. The samples are then separated via high-performance liquid chromatography (HPLC) fractionation, and the fractions are treated with 2,4,6-trinitrobenzenesulfonic acid (TNBS). The TNBS reacts with

the free amine of tryptic peptides, adding a trinitrophenyl group to tryptic peptides and increasing their hydrophobicity. Finally, the acetyl-labeled native and neo-N-terminal peptides and TNBS labeled tryptic peptides mixture are separated by the secondary HPLC run. Since TNBS labeling alters the hydrophobicity of tryptic peptides, they run at a different retention time during the secondary HPLC run, while the retention time of acetyl-labeled native and neo-N-terminal peptides remains the same during primary and secondary HPLC runs, enabling separation of tryptic peptides. These native and neo-N-terminal peptides are then collected and analyzed using LC-MS/MS analysis [78]. LC-MS/MS analysis of acetyl-labeled native or neo-N-terminal peptides provide the identity of protease substrates and the protease cleavage sites (Figure 3A). Despite having many advantages, COFRADIC also has certain limitations. Apart from being an expensive technique, sample preparation and data analysis is labor-intensive and require significant expertise. In addition, since the method includes multiple chromatographic runs, it requires a large number of samples [79].

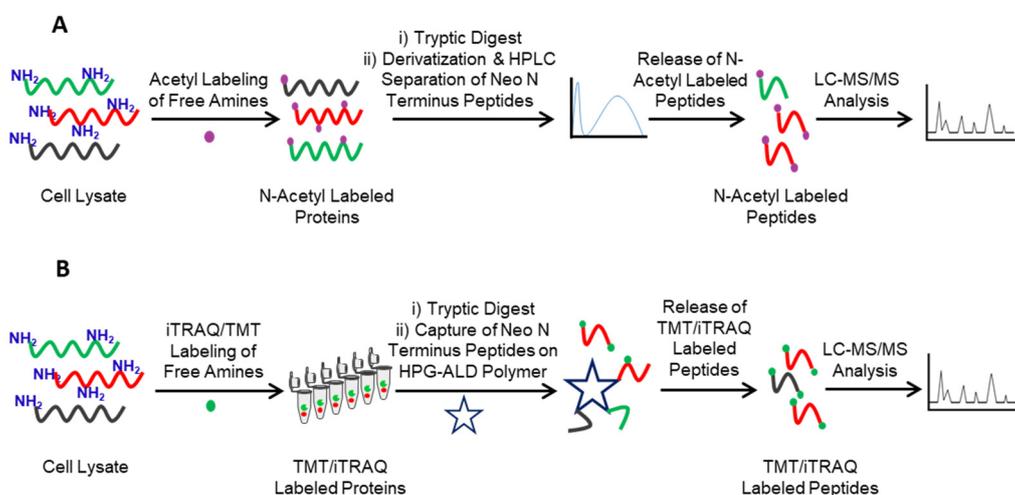


Figure 3. Negative enrichment approaches of N-terminomics. **(A)** In COFRADIC, the native or neo-N-terminal peptides are blocked by acetylation followed by tryptic digest. The internal tryptic peptides are then labeled and separated by HPLC, leaving acetyl-labeled native or neo-N-terminal peptides. LC-MS/MS analysis of acetyl-labeled native and neo-N-terminal peptides provide identity of protease substrates and protease cleavage sites. **(B)** In TAILS, the native or neo-N-terminal peptides are blocked with isobaric labeling followed by trypsin digestion. The internal tryptic peptides are then separated by capturing HPG-ALD polymer, leaving isobaric tag labeled native and neo-N-terminal peptides. LC-MS/MS analysis of probe-labeled native and neo-N-terminal peptides provide identity of protease substrates and protease cleavage sites.

Zahedi and co-workers recently modified COFRADIC to charge-based fractional diagonal chromatography (ChaFRADIC) [80,81]. Instead of using the modified hydrophobicity of tryptic peptides for chromatographic separation, this method uses the altered charged state of tryptic peptides to separate them from native or neo-N-terminal peptides. In this method, native or neo-N-terminal peptides are blocked by carbamidomethylation followed by trypsin digestion. The samples are then separated by a strong cation exchange (SCX) chromatography at pH 2.7 in different fractions, and the tryptic peptides in these fractions are blocked by acetylation, altering the charge of the tryptic peptides [82]. Finally, the carbamidomethyl-labeled native and neo-N-terminal peptides and acetyl-labeled tryptic peptides mixture are separated by the secondary SCX run. Since acetylation alters the charge of tryptic peptides, they run at a different retention time during the secondary SCX run, while the retention time of carbamidomethylated native and neo-N-terminal peptides remains the same during primary and secondary SCX runs, enabling separation of tryptic peptides. These carbamidomethyl-labeled native and neo-N-terminal peptides are then collected and analyzed using LC-MS/MS analysis [81]. LC-MS/MS analysis of native or neo-N-terminal peptides provides the identity of protease substrates and the protease

cleavage sites (Figure 3A). The main advantages of ChaFRADIC are reductions in both the amount of material required and LC-MS/MS instrument time.

The COFRADIC method is widely used to study the backbone N-terminal acetylation in eukaryotes [32]. COFRADIC revealed that about 84% of human proteins and 57% of yeast proteins were acetylated at the N-terminus. The authors further noted that about 50% of the acetylation occurred on a protein containing a Met-Lys dipeptide sequence in humans, a phenomenon not observed in the yeast [44]. COFRADIC has also been used to construct the substrate degradome of apoptosis by identifying caspase substrates in human Jurkat T-lymphocytes. Using COFRADIC, the authors identified 14 different caspase substrates involved in the spliceosome, suggesting the importance of alternative splicing in apoptosis [45,83]. Further studies revealed that the initiator caspase 2 and executioner caspases 3 and 7 share a common consensus sequence DEVD|G [46]. Additionally, Impens et al. used the COFRADIC method to identify human immunodeficiency virus-1 (HIV-1) protease substrates, showing the prominent role of HIV proteases in HIV infection [47].

TAILS Terminal Amine Isotopic Labeling of Substrates (TAILS) an alternative negative enrichment method developed by Christopher Overall and co-workers to overcome the limitations of COFRADIC [84]. In contrast to COFRADIC, in TAILS, the internal tryptic peptides are separated using a primary amine-reactive polymer which considerably reduces the time and effort required for the separation (Figure 3B). The basic TAILS protocol starts with blocking the native and neo-N-terminal peptides using isotopic reagents such as Tandem Mass Tagging (TMT) or iTRAQ, followed by trypsin digestion. The internal tryptic peptides are then captured using high molecular weight dendritic polyglycerol aldehyde polymer (HPG-ALD). The HPG-ALD polymer covalently reacts with the amino terminus of tryptic peptides, leaving the isotopically labeled native and neo-N-terminal peptides unbound which are then recovered by ultrafiltration and analyzed using LC-MS/MS analysis. LC-MS/MS analysis of isobaric tag labeled native or neo-N-terminal peptides provide identity protease substrates and the protease cleavage sites (Figure 3B) [85]. Alternatively, Mommen et al. developed a phospho tagging (PTAG) method that can also be used to label the internal tryptic peptides, which are then depleted using titanium dioxide (TiO₂), leaving the isotopically labeled native and neo-N-termini unbound, which are collected and analyzed using LC-MS/MS analysis. LC-MS/MS analysis of isobaric tag labeled native and neo-N-terminal peptides provide identification of protease substrates and the protease cleavage sites [86].

TAILS possesses several advantages over the COFRADIC approach, such as N-terminal specificity biases are minimal with this approach, a smaller quantity of starting material is required, and kits and reagents are available commercially [79]. However, despite having many advantages, TAILS also has some bottlenecks. For example, TAILS requires significant statistical analysis to distinguish the tagged and untagged N-termini in complex peptide mixtures, and the reagents used in this method are very expensive. Furthermore, the negative enrichment methods, COFRADIC and TAILS, could suffer from high false positives due to the incomplete blocking and isolation of the internal tryptic peptides [87].

Since its development, the TAILS method has been successfully applied to identify protease substrates from many different systems. Using TAILS, Klein et al. identified HOIL1 as a substrate of the paracaspase, mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1). In the presence of antigens, HOLT1 amplifies the immune response during the NF- κ B mediated signaling [67]. Using TAILS, Gordon et al. compared the relative proteolysis between the ulcerative colitis patients and healthy controls. The authors noted that the host proteases were the major contributors to overall proteolysis. These studies further showed that proteolysis of proteins involved in neutrophil degranulation, metabolism, and adherens junction was altered in ulcerative colitis patients [65]. Additionally, TAILS has been used to identify the proteolytic profile of chronic obstructive pulmonary disease (COPD) samples. These studies identified a total of 299 human sputum proteins, out of which 125 were substrates of proteases [66].

TAILS has been widely used to identify the novel physiological substrates of metallo and other proteases. Using iTRAQ-TAILS, Padova et al. identified the substrate profile of substrates of MMP-2 and MMP-9. These studies confirmed that insulin-like growth factor-binding protein-4, thrombospondin-2, and galectin-1 are substrates of MMP-2, and MMP-9, establishing the role of their proteases in angiogenesis [41]. Moreover, Starr et al. identified the substrate profile of a neutrophil-specific membrane-type 6 matrix metalloproteinase (MTP6-MMP), which plays a role in cancer and multiple sclerosis. These studies established vimentin, cystatin C, secreted protein, acidic and rich in cysteine (SPARC), and galectin-1 as substrates of MTP6-MMP [71]. Bellac et al. used TAILS to identify the substrate profile of a macrophage-specific metalloproteinase MMP12 involved in inflammation. Authors showed that MMP-12 activates prothrombin and reduces inflammation by decreasing complement activation [72]. TAILS was also used to identify the substrate profile of metalloprotease meprin β . These studies confirm that meprin β cleaves the amyloid precursor protein, a critically important protein in neurodegeneration [73]. Moreover, Alcaraz et al. used TAILS to identify the substrates of a tumor-specific protease, cathepsin D, from triple-negative breast cancer cells. The authors reported that Cathepsin D cleaves the matricellular SPARC protein under acidic pH, generating a 9-KDa fragment, which is more oncogenic than intact SPARC [64]. TAILS was also used to identify substrates of the 3C proteinases (3C^{Pro}) of coxsackievirus B3 and poliovirus. These studies showed that the heterogeneous nuclear ribonucleoprotein M (hnRNP M), an RNA-binding protein that plays a role in pre-mRNA splicing, is a substrate of 3C^{Pro} [75].

Subtiligase: In contrast to COFRADIC and TAILS, Subtiligase is a positive N-terminomics method that directly enriches and isolates native or neo-N-terminal peptides. The Subtiligase method is developed by Jim Wells and co-workers, which uses the enzymatic biotinylation of the native or neo-N-terminal peptides using a modified Subtiligase enzyme (Figure 4A) [88]. The labeled peptides are then enriched, isolated, and identified using LC-MS/MS analysis. Subtiligase is engineered from a nonspecific protease called subtilisin to selectively modify N-terminal amines while leaving the protein ϵ -lysine side chains unmodified. In the Subtiligase method, the native or neo-N-terminal proteins/peptides are labeled with the TEV-biotin tag using subtiligase. The probe-labeled native or neo-N-terminal peptides are then enriched on streptavidin beads, followed by sequential on-bead trypsin and Tobacco Etch Virus (TEV) digestion and elution of the probe-labeled native and neo-N-terminal peptides. The TEV-cleaved neo-N-terminal peptides consist of an α -aminobutyric acid signature, differentiating them from tryptic peptides [32]. LC-MS/MS analysis of the probe-labeled native or neo-N-terminal peptides provides identity of protease substrates and protease cleavage sites (Figure 4A). Recently, Jim Wells and co-workers further optimized the Subtiligase method to identify protease cleavage sites from the membrane proteins present on the surface of living cells, making it the only N-terminomics method that can be used to identify protease substrates from the living cells [56].

The Subtiligase approach holds many advantages over the other N-terminomics methods, such as live-cell labeling, direct enrichment, and explicit identification of neo-N-termini without ambiguity. However, despite many advantages, there are a few drawbacks, such as the starting large quantities of starting material is required for adequate enrichment, and the reagents used in this approach are not commercially available. In addition, although Subtiligase has a higher and broader specificity than other ligases, it still has biases towards specific N-terminal sequences.

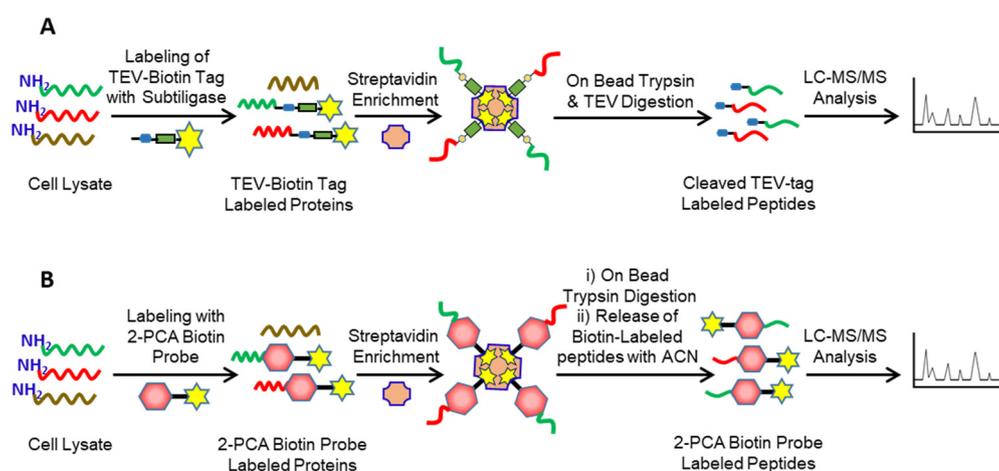


Figure 4. Positive enrichment approaches of N-terminomics. (A) In Subtiligase approach, the native or neo-N-terminal peptides labeled with the TEV-biotin probe using Subtiligase enzyme. The labeled peptides are then enriched on avidin beads followed by sequential on-bead trypsin and TEV digestion. LC-MS/MS analysis of probe-labeled native and neo-N-terminal peptides provides the identity of protease substrates and protease cleavage sites. (B) In CHOPS, the native or neo-N-terminal peptides are labeled with the 2PCA-biotin tag. The probe-labeled peptides are then enriched on avidin beads followed by on-bead trypsin digestion. After washing, the probe-labeled native and neo-N-terminal peptides are eluted from the beads using acetonitrile. LC-MS/MS analysis of probe-labeled native and neo-N-terminal peptides provides the identity of protease substrates and protease cleavage sites.

Subtiligase has been widely adopted to identify the global proteolytic events and targeted protease substrate specificity from biological samples using *forward* and *reverse* N-terminomics approaches. The *forward* Subtiligase N-terminomics was used to identify the proteolytic profile of human blood and plasma samples. In this study, the authors identified multiple N-terminal peptides originating from complement and coagulation cascades. Additionally, the *forward* Subtiligase was used to identify the global proteomic profile of thyroid carcinoma, colorectal cancer, Alzheimer’s disease, and cardiovascular diseases [57,59–63]. Overall, the Subtiligase method enabled the simplification of human blood proteomes by specifically enriching the N-terminal peptides and discovering disease-specific biomarkers.

The *reverse* Subtiligase N-terminomics has explicitly been used to study the substrate specificity of caspases, a class of cysteine proteases involved in apoptosis and pyroptosis, including caspases 3, 6, 7 (executioner caspases) [53,54,89], and caspases 1, 4, and 5 (inflammatory caspase) [48,49,90]. Caspases cleave their substrates predominantly after aspartate residues [50,51]. Using the *reverse* Subtiligase approach, Scaman et al. showed that apart from aspartic acid, caspases 3 and 7 can cleave after glutamic acid residues, while caspase 3 can also cleave specifically after phosphorylated serine residues [52]. In addition, using Subtiligase, Agard et al. identified about 82 caspase-1 and three caspase-4 substrates from THP-1 monocytic cell lysates. The authors further showed that caspase-1 contains a higher number of cellular substrates than caspases-4 and 5 [90]. Additionally, using the *reverse* Subtiligase approach, Hill et al. identified the substrate profile of Zika virus protease NS2B-NS3 from HEK293 cells. The authors showed that the Zika virus protease cleaves host eukaryotic translation initiation factor 4 gamma 1 (eIF4G1) and autophagy-related protein 16-1 (ATG16L1) [55].

CHOPS: Chemical enrichment of Protease Substrates (CHOPS) is a positive enrichment method recently developed by Daniel Bachovchin and co-workers as an alternative approach to Subtiligase [76]. In contrast to Subtiligase, CHOPS uses the chemical biotinylation of the native or neo-N-terminal peptides using a 2-pyridinecarboxaldehyde (2PCA)-biotin probe. The probe-labeled peptides are then enriched, isolated, and identified using LC-MS/MS analysis (Figure 4B). 2PCA selectively modifies N-terminal amines while leaving the protein ϵ -lysine side chains unmodified. In the CHOPS method, the native and

neo-N-terminal proteins/peptides are labeled with the 2PCA-biotin tag. The probe-labeled are then enriched on streptavidin beads, followed by on-bead trypsin digestion. After removing tryptic peptides, the probe-labeled native and neo-N-terminal peptides are eluted from the beads using acetonitrile followed by LC-MS/MS analysis. LC-MS/MS analysis of probe-labeled native and neo-N-terminal peptides provides the identity of protease substrates and the protease cleavage sites (Figure 4B).

Since CHOPS uses chemical biotinylation instead of enzymatic biotinylation, it removed the major flaw of the Subtiligase method, i.e., biases in labeling towards N-terminal sequences, while still retaining all advantages of the Subtiligase method, including direct enrichment and explicit identification of neo-N-termini without ambiguity. However, this method still needs to be optimized for live-cell labeling. Nonetheless, CHOPS remains the most versatile and promising N-terminomics method for future use.

CHOPS was used to identify substrate specificity of dipeptidyl peptidases (DPP). The dipeptidyl peptidases (DPPs) cleave after proline residues present on the second position in polypeptide chains, releasing dipeptides from the N-termini. The DPPs regulate the activities of many biologically important molecules such as neuropeptides, cytokines, and hormones [91]. However, the identification of the DPP substrate remains challenging predominantly due to the similar size of the intact and cleaved substrates [92,93]. Griswold et al. successfully employed the CHOPS method to identify the substrate specificity of DPP8 and DPP9. The authors showed that DPP8 and DPP9 do not cleave full-length globular proteins Nlrp1b and SMAC but efficiently cleave short peptides, confirming that DPP8 and DPP9 predominately prefer small peptide substrates over full length proteins [76].

3. Conclusions and Future Directions

Identifying protease substrates and proteolytic events under physiological conditions is a fundamental step towards understanding the proteolytic landscape of biological samples. The N-terminomics methods discussed in this review are emerging as indispensable tools to identify and understand the global proteolytic events and protease substrate specificity under physiological conditions. The main steps in N-terminomics methods are the labeling and enrichment of the native N-termini or neo-N termini peptides generated from the proteolytic cleavage, reducing the complexity of the samples, and enabling the unambiguous identification of the protease substrates and protease cleavage sites.

In recent years, multiple positive and negative enrichment N-terminomics strategies have been developed. Although the N-terminomics field is rapidly expanding, few challenges are associated with these techniques, requiring further technological and computational progression. For example, the precise annotation of the protease cleavage sites still remains challenging as they are annotated based on protein backtracking from the single peptide hit. In addition, some unfavorably cleaved neo-N-terminal peptides cannot be detected during LC-MS/MS analysis. Furthermore, the regulation of proteases by post-translational modifications remains another challenging avenue to study by N-terminomics as these post-translational modifications can lead to a generation of non-tryptic or semi-tryptic peptides and can be difficult to detect. Such challenges can be overcome by an integrated approach, including both neo-N-terminal and neo-C-terminal peptides analysis. The simultaneous analysis of N- and C-terminomics could improve the annotation of cleavage sites and provide much more information about the global proteolytic events in complex biological mixtures.

The current N-terminomics strategies can be further used in several different directions to increase our knowledge of the complex interactions between proteases, their substrates, and inhibitors. For example, N-terminomics can be used to identify and annotate the unpredicted and undefined proteins and proteoforms to improve the existing genomic databases. Moreover, the *reverse* N-terminomics can be increasingly used to map the protein–protease interactome of the target proteases under physiological conditions. Such data can then be used to dissect the protease regulation in a proteolytic cascade and generate a proteolytic network in health and diseases. In addition, such data can help

to develop a library of proteolytic biomarkers that could be used to diagnose multiple pathological conditions.

In conclusion, N-terminomics methods have opened a new horizon to study protease regulation from the complex biological mixtures, far from our reach a decade ago. More studies are underway using these techniques to decode the biological protease pathways and understand the crosstalk between proteolysis and post-translational modifications. Such studies will generate a future roadmap for identifying novel disease mechanisms, drug targets, and biomarkers for multiple pathologies.

Funding: This research received no external funding.

Acknowledgments: This work was supported in part by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under award number P20GM113117.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Combinatorial Fractional Diagonal Chromatography (COFRADIC), Terminal Amine Isotopic Labeling of Substrates (TAILS), Chemical enrichment of Protease Substrates (CHOPS), Tobacco Etch Virus protease (TEV protease), dipeptidyl peptidases (DPP), Matrix Metallo Protease (MMP), polyglycerol aldehyde polymer (HPG-ALD), Tandem Mass Tagging (TMT), strong cation exchange chromatography (SCX).

References

1. Gurumalles, P.; Alagu, K.; Ramakrishnan, B.; Muthusamy, S. A systematic reconsideration on proteases. *Int. J. Biol. Macromol.* **2019**, *128*, 254–267. [[CrossRef](#)]
2. Quesada, V.; Ordóñez, G.R.; Sánchez, L.M.; Puente, X.S.; López-Otín, C. The Degradome database: Mammalian proteases and diseases of proteolysis. *Nucleic Acids Res.* **2008**, *37* (Suppl. 1), D239–D243. [[CrossRef](#)] [[PubMed](#)]
3. Rawlings, N.D.; Morton, F.R.; Barrett, A.J. MEROPS: The peptidase database. *Nucleic Acids Res.* **2006**, *34* (Suppl. 1), D270–D272. [[CrossRef](#)] [[PubMed](#)]
4. López-Otín, C.; Bond, J.S. Proteases: Multifunctional enzymes in life and disease. *J. Biol. Chem.* **2008**, *283*, 30433–30437. [[CrossRef](#)]
5. King, R.W.; Deshaies, R.J.; Peters, J.M.; Kirschner, M.W. How proteolysis drives the cell cycle. *Science* **1996**, *274*, 1652–1659. [[CrossRef](#)] [[PubMed](#)]
6. Scher, W. The role of extracellular proteases in cell proliferation and differentiation. *Lab. Investig. J. Tech. Methods Pathol.* **1987**, *57*, 607–633.
7. Sukharev, S.A.; Pleshakova, O.V.; Sadovnikov, V.B. Role of proteases in activation of apoptosis. *Cell Death Differ.* **1997**, *4*, 457–462. [[CrossRef](#)]
8. Ruggiano, A.; Ramadan, K. DNA–protein crosslink proteases in genome stability. *Commun. Biol.* **2021**, *4*, 11. [[CrossRef](#)] [[PubMed](#)]
9. Chen, W.T. Membrane proteases: Roles in tissue remodeling and tumour invasion. *Curr. Opin. Cell Biol.* **1992**, *4*, 802–809. [[CrossRef](#)]
10. Walsh, P.N.; Ahmad, S.S. Proteases in blood clotting. *Essays Biochem.* **2002**, *38*, 95–111.
11. Singh, N.; Bhattacharyya, D. Proteases in Wound Healing and Immunity. In *Proteases in Human Diseases*; Chakraborti, S., Chakraborti, T., Dhalla, N.S., Eds.; Springer: Singapore, 2017; pp. 147–170.
12. Eatemadi, A.; Aiyelabegan, H.T.; Negahdari, B.; Mazlomi, M.A.; Daraee, H.; Daraee, N.; Eatemadi, R.; Sadroddiny, E. Role of protease and protease inhibitors in cancer pathogenesis and treatment. *Biomed. Pharmacother.* **2017**, *86*, 221–231. [[CrossRef](#)]
13. Hasanbasic, S.; Jahic, A.; Karahmet, E.; Sejrancic, A.; Prnjavorac, B. The role of cysteine protease in Alzheimer disease. *Mater. Socio-Med.* **2016**, *28*, 235. [[CrossRef](#)]
14. Suchiva, P.; Takai, T.; Kamijo, S.; Maruyama, N.; Yokomizo, T.; Sugimoto, Y.; Okumura, K.; Ikeda, S.; Ogawa, H. Inhibition of Both Cyclooxygenase-1 and-2 Promotes Epicutaneous Th2 and Th17 Sensitization and Allergic Airway Inflammation on Subsequent Airway Exposure to Protease Allergen in Mice. *Int. Arch. Allergy Immunol.* **2021**, 1–12. [[CrossRef](#)] [[PubMed](#)]
15. Bernardi, F.; Mariani, G. Biochemical, molecular and clinical aspects of coagulation factor VII and its role in hemostasis and thrombosis. *Haematologica* **2021**, *106*, 351. [[CrossRef](#)] [[PubMed](#)]
16. Hua, Y.; Nair, S. Proteases in cardiometabolic diseases: Pathophysiology, molecular mechanisms and clinical applications. *Biochim. Biophys. Acta* **2015**, *1852*, 195–208. [[CrossRef](#)] [[PubMed](#)]
17. Overall, C.M.; Lopez-Otin, C. Strategies for MMP inhibition in cancer: Innovations for the post-trial era. *Nat. Rev. Cancer* **2002**, *2*, 657–672. [[CrossRef](#)] [[PubMed](#)]
18. Yan, C.; Boyd, D.D. Regulation of matrix metalloproteinase gene expression. *J. Cell. Physiol.* **2007**, *211*, 19–26. [[CrossRef](#)] [[PubMed](#)]
19. Riedl, S.J.; Salvesen, G.S. The apoptosome: Signalling platform of cell death. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 405–413. [[CrossRef](#)]

20. Versteeg, H.H.; Ruf, W. Emerging insights in tissue factor-dependent signaling events. *Semin. Thromb. Hemost.* **2006**, *32*, 24–32. [[CrossRef](#)] [[PubMed](#)]
21. Fu, X.; Parks, W.C.; Heinecke, J.W. Activation and silencing of matrix metalloproteinases. *Semin. Cell Dev. Biol.* **2008**, *19*, 2–13. [[CrossRef](#)]
22. Bode, W.; Huber, R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim. Biophys. Acta* **2000**, *1477*, 241–252. [[CrossRef](#)]
23. Turk, B. Targeting proteases: Successes, failures and future prospects. *Nat. Rev. Drug Discov.* **2006**, *5*, 785–799. [[CrossRef](#)]
24. Mo, X.Y.; Cascio, P.; Lemerise, K.; Goldberg, A.L.; Rock, K. Distinct Proteolytic Processes Generate the C and N Termini of MHC Class I-Binding Peptides. *J. Immunol.* **1999**, *163*, 5851–5859.
25. Khan, A.R.; James, M.N. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **1998**, *7*, 815–836. [[CrossRef](#)] [[PubMed](#)]
26. Voisin, A.; Monville, C.; Plancheron, A.; Béré, E.; Gaillard, A.; Levezuel, N. Cathepsin B pH-Dependent Activity Is Involved in Lysosomal Dysregulation in Atrophic Age-Related Macular Degeneration. *Oxid. Med. Cell. Longev.* **2019**, *2019*, 5637075. [[CrossRef](#)] [[PubMed](#)]
27. Wang, R.; Edrington, T.C.; Storrs, S.B.; Crowley, K.S.; Ward, J.M.; Lee, T.C.; Liu, Z.L.; Li, B.; Glenn, K.C. Analyzing pepsin degradation assay conditions used for allergenicity assessments to ensure that pepsin susceptible and pepsin resistant dietary proteins are distinguishable. *PLoS ONE* **2017**, *12*, e0171926. [[CrossRef](#)]
28. Allgayer, H.; Lengyel, E.; Boyd, D.D. Transcriptional Control of Proteases. In *Proteases and Their Inhibitors in Cancer Metastasis*; Foidart, J.-M., Muschel, R.J., Eds.; Springer: Dordrecht, The Netherlands, 2002; pp. 151–168.
29. Klein, T.; Eckhard, U.; Dufour, A.; Solis, N.; Overall, C.M. Proteolytic Cleavage Mechanisms, Function, and “Omic” Approaches for a Near-Ubiquitous Posttranslational Modification. *Chem. Rev.* **2018**, *118*, 1137–1168. [[CrossRef](#)] [[PubMed](#)]
30. Waugh, D.S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **2011**, *80*, 283–293. [[CrossRef](#)]
31. Varshavsky, A. The N-end rule pathway of protein degradation. *Genes Cells* **1997**, *2*, 13–28. [[CrossRef](#)] [[PubMed](#)]
32. Luo, S.Y.; Araya, L.E.; Julien, O. Protease substrate identification using N-terminomics. *ACS Chem. Biol.* **2019**, *14*, 2361–2371. [[CrossRef](#)]
33. Angel, T.E.; Aryal, U.K.; Hengel, S.M.; Baker, E.S.; Kelly, R.T.; Robinson, E.W.; Smith, R.D. Mass spectrometry-based proteomics: Existing capabilities and future directions. *Chem. Soc. Rev.* **2012**, *41*, 3912–3928. [[CrossRef](#)]
34. Timmer, J.C.; Salvesen, G.S. N-terminomics: A high-content screen for protease substrates and their cleavage sites. *Methods Mol. Biol.* **2011**, *753*, 243–255.
35. Tanco, S.; Gevaert, K.; Van Damme, P. C-terminomics: Targeted analysis of natural and posttranslationally modified protein and peptide C-termini. *Proteomics* **2015**, *15*, 903–914. [[CrossRef](#)] [[PubMed](#)]
36. Wiita, A.P.; Seaman, J.E.; Wells, J.A. Global analysis of cellular proteolysis by selective enzymatic labeling of protein N-termini. *Methods Enzymol.* **2014**, *544*, 327–358. [[PubMed](#)]
37. Löffek, S.; Schilling, O.; Franzke, C.W. Biological role of matrix metalloproteinases: A critical balance. *Eur. Respir. J.* **2011**, *38*, 191–208. [[CrossRef](#)]
38. Holmberg, C.; Ghesquière, B.; Impens, F.; Gevaert, K.; Kumar, J.D.; Cash, N.; Kandola, S.; Hegyi, P.; Wang, T.C.; Dockray, G.J.; et al. Mapping proteolytic processing in the secretome of gastric cancer-associated myofibroblasts reveals activation of MMP-1, MMP-2, and MMP-3. *J. Proteome Res.* **2013**, *12*, 3413–3422. [[CrossRef](#)]
39. Prudova, A.; Gocheva, V.; Auf dem Keller, U.; Eckhard, U.; Olson, O.C.; Akkari, L.; Butler, G.S.; Fortelny, N.; Lange, P.F.; Mark, J.C. TAILS N-terminomics and proteomics show protein degradation dominates over proteolytic processing by cathepsins in pancreatic tumors. *Cell Rep.* **2016**, *16*, 1762–1773. [[CrossRef](#)] [[PubMed](#)]
40. Colige, A.; Monseur, C.; Crawley, J.T.B.; Santamaria, S.; de Groot, R. Proteomic discovery of substrates of the cardiovascular protease ADAMTS7. *J. Biol. Chem.* **2019**, *294*, 8037–8045. [[CrossRef](#)]
41. Prudova, A.; Auf dem Keller, U.; Butler, G.S.; Overall, C.M. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol. Cell. Proteom.* **2010**, *9*, 894–911. [[CrossRef](#)]
42. Plasman, K.; Demol, H.; Bird, P.I.; Gevaert, K.; Van Damme, P. Substrate specificities of the granzyme tryptases A and K. *J. Proteome Res.* **2014**, *13*, 6067–6077. [[CrossRef](#)]
43. Crawford, E.D.; Wells, J.A. Caspase substrates and cellular remodeling. *Annu. Rev. Biochem.* **2011**, *80*, 1055–1087. [[CrossRef](#)]
44. Arnesen, T.; Van Damme, P.; Polevoda, B.; Helsens, K.; Evjenth, R.; Colaert, N.; Varhaug, J.E.; Vandekerckhove, J.; Lillehaug, J.R.; Sherman, F.; et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8157–8162. [[CrossRef](#)]
45. Van Damme, P.; Martens, L.; Van Damme, J.; Hugelier, K.; Staes, A.; Vandekerckhove, J.; Gevaert, K. Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat. Methods* **2005**, *2*, 771–777. [[CrossRef](#)] [[PubMed](#)]
46. Wejda, M.; Impens, F.; Takahashi, N.; Van Damme, P.; Gevaert, K.; Vandenebeele, P. Degradomics reveals that cleavage specificity profiles of caspase-2 and effector caspases are alike. *J. Biol. Chem.* **2012**, *287*, 33983–33995. [[CrossRef](#)] [[PubMed](#)]
47. Impens, F.; Timmerman, E.; Staes, A.; Moens, K.; Arien, K.K.; Verhasselt, B.; Vandekerckhove, J.; Gevaert, K. A catalogue of putative HIV-1 protease host cell substrates. *Biol. Chem.* **2012**, *393*, 915–931. [[CrossRef](#)] [[PubMed](#)]
48. Lamkanfi, M.; Declercq, W.; Kalai, M.; Saelens, X.; Vandenebeele, P. Alice in caspase land. A phylogenetic analysis of caspases from worm to man. *Cell Death Differ.* **2002**, *9*, 358–361. [[CrossRef](#)] [[PubMed](#)]

49. Thornberry, N.A.; Lazebnik, Y. Caspases: Enemies within. *Science* **1998**, *281*, 1312–1316. [[CrossRef](#)]
50. Thornberry, N.A.; Rano, T.A.; Peterson, E.P.; Rasper, D.M.; Timkey, T.; Garcia-Calvo, M.; Houtzager, V.M.; Nordstrom, P.A.; Roy, S.; Vaillancourt, J.P.; et al. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J. Biol. Chem.* **1997**, *272*, 17907–17911. [[CrossRef](#)]
51. Stennicke, H.R.; Renatus, M.; Meldal, M.; Salvesen, G.S. Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8. *Biochem. J.* **2000**, *350 Pt 2*, 563–568. [[CrossRef](#)]
52. Seaman, J.E.; Julien, O.; Lee, P.S.; Rettenmaier, T.J.; Thomsen, N.D.; Wells, J.A. Cacidases: Caspases can cleave after aspartate, glutamate and phosphoserine residues. *Cell Death Differ.* **2016**, *23*, 1717–1726. [[CrossRef](#)]
53. Agard, N.J.; Mahrus, S.; Trinidad, J.C.; Lynn, A.; Burlingame, A.L.; Wells, J.A. Global kinetic analysis of proteolysis via quantitative targeted proteomics. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1913–1918. [[CrossRef](#)] [[PubMed](#)]
54. Hill, M.E.; MacPherson, D.J.; Wu, P.; Julien, O.; Wells, J.A.; Hardy, J.A. Reprogramming Caspase-7 Specificity by Regio-Specific Mutations and Selection Provides Alternate Solutions for Substrate Recognition. *ACS Chem. Biol.* **2016**, *11*, 1603–1612. [[CrossRef](#)]
55. Hill, M.E.; Kumar, A.; Wells, J.A.; Hobman, T.C.; Julien, O.; Hardy, J.A. The Unique Cofactor Region of Zika Virus NS2B-NS3 Protease Facilitates Cleavage of Key Host Proteins. *ACS Chem. Biol.* **2018**, *13*, 2398–2405. [[CrossRef](#)] [[PubMed](#)]
56. Weeks, A.M.; Byrnes, J.R.; Lui, I.; Wells, J.A. Mapping proteolytic neo-N termini at the surface of living cells. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018809118. [[CrossRef](#)]
57. Wildes, D.; Wells, J.A. Sampling the N-terminal proteome of human blood. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4561–4566. [[CrossRef](#)]
58. Mizuguchi, M.; Hayashi, A.; Takeuchi, M.; Dobashi, M.; Mori, Y.; Shinoda, H.; Aizawa, T.; Demura, M.; Kawano, K. Unfolding and aggregation of transthyretin by the truncation of 50 N-terminal amino acids. *Proteins* **2008**, *72*, 261–269. [[CrossRef](#)] [[PubMed](#)]
59. Janowski, R.; Abrahamson, M.; Grubb, A.; Jaskolski, M. Domain swapping in N-truncated human cystatin C. *J. Mol. Biol.* **2004**, *341*, 151–160. [[CrossRef](#)]
60. Liz, M.A.; Gomes, C.M.; Saraiva, M.J.; Sousa, M.M. ApoA-I cleaved by transthyretin has reduced ability to promote cholesterol efflux and increased amyloidogenicity. *J. Lipid Res.* **2007**, *48*, 2385–2395. [[CrossRef](#)] [[PubMed](#)]
61. Villanueva, J.; Nazarian, A.; Lawlor, K.; Robbins, R.J.; Tempst, P. A sequence-specific exopeptidase activity test (SSEAT) for “functional” biomarker discovery. *Mol. Cell. Proteom.* **2008**, *7*, 509–518. [[CrossRef](#)]
62. Fuchs, C.S.; Goldberg, R.M.; Sargent, D.J.; Meyerhardt, J.A.; Wolpin, B.M.; Green, E.M.; Pitot, H.C.; Pollak, M. Plasma insulin-like growth factors, insulin-like binding protein-3, and outcome in metastatic colorectal cancer: Results from intergroup trial N9741. *Clin. Cancer Res.* **2008**, *14*, 8263–8269. [[CrossRef](#)] [[PubMed](#)]
63. Harris, F.M.; Brecht, W.J.; Xu, Q.; Tesseur, I.; Kekoni, L.; Wyss-Coray, T.; Fish, J.D.; Masliah, E.; Hopkins, P.C.; Scarse-Levie, K. Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer’s disease-like neurodegeneration and behavioral deficits in transgenic mice. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10966–10971. [[CrossRef](#)]
64. Alcaraz, L.B.; Mallavialle, A.; David, T.; Derocq, D.; Delolme, F.; Dieryckx, C.; Mollevi, C.; Boissiere-Michot, F.; Simony-Lafontaine, J.; Du Manoir, S.; et al. A 9-kDa matricellular SPARC fragment released by cathepsin D exhibits pro-tumor activity in the triple-negative breast cancer microenvironment. *Theranostics* **2021**, *11*, 6173–6192. [[CrossRef](#)]
65. Gordon, M.H.; Anowai, A.; Young, D.; Das, N.; Campden, R.I.; Sekhon, H.; Myers, Z.; Mainoli, B.; Chopra, S.; Thuy-Boun, P.S.; et al. N-Terminomics/TAILS Profiling of Proteases and Their Substrates in Ulcerative Colitis. *ACS Chem. Biol.* **2019**, *14*, 2471–2483. [[CrossRef](#)] [[PubMed](#)]
66. Mallia-Milanes, B.; Dufour, A.; Philp, C.; Solis, N.; Klein, T.; Fischer, M.; Bolton, C.E.; Shapiro, S.; Overall, C.M.; Johnson, S.R. TAILS proteomics reveals dynamic changes in airway proteolysis controlling protease activity and innate immunity during COPD exacerbations. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **2018**, *315*, L1003–L1014. [[CrossRef](#)] [[PubMed](#)]
67. Klein, T.; Fung, S.Y.; Renner, F.; Blank, M.A.; Dufour, A.; Kang, S.; Bolger-Munro, M.; Scurll, J.M.; Priatel, J.J.; Schweigler, P.; et al. The paracaspase MALT1 cleaves HOIL1 reducing linear ubiquitination by LUBAC to dampen lymphocyte NF-kappaB signalling. *Nat. Commun.* **2015**, *6*, 8777. [[CrossRef](#)] [[PubMed](#)]
68. Wilson, C.H.; Indarto, D.; Doucet, A.; Pogson, L.D.; Pitman, M.R.; McNicholas, K.; Menz, R.I.; Overall, C.M.; Abbott, C.A. Identifying Natural Substrates for Dipeptidyl Peptidases 8 and 9 Using Terminal Amine Isotopic Labeling of Substrates (TAILS) Reveals in Vivo Roles in Cellular Homeostasis and Energy Metabolism. *J. Biol. Chem.* **2013**, *288*, 13936–13949. [[CrossRef](#)] [[PubMed](#)]
69. Jefferson, T.; Auf dem Keller, U.; Bellac, C.; Metz, V.V.; Broder, C.; Hedrich, J.; Ohler, A.; Maier, W.; Magdolen, V.; Sterchi, E. The substrate degradome of meprin metalloproteases reveals an unexpected proteolytic link between meprin β and ADAM10. *Cell. Mol. Life Sci.* **2013**, *70*, 309–333. [[CrossRef](#)]
70. Zelanis, A.; Oliveira, A.K.; Prudova, A.; Huesgen, P.F.; Tashima, A.K.; Kizhakkedathu, J.; Overall, C.M.; Serrano, S.M. Deep profiling of the cleavage specificity and human substrates of snake venom metalloprotease HF3 by proteomic identification of cleavage site specificity (PICS) using proteome derived peptide libraries and terminal amine isotopic labeling of substrates (TAILS) N-terminomics. *J. Proteome Res.* **2019**, *18*, 3419–3428.
71. Starr, A.E.; Bellac, C.L.; Dufour, A.; Goebeler, V.; Overall, C.M. Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): Chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J. Biol. Chem.* **2012**, *287*, 13382–13395. [[CrossRef](#)]

72. Bellac, C.L.; Dufour, A.; Krisinger, M.J.; Loonchanta, A.; Starr, A.E.; Auf dem Keller, U.; Lange, P.F.; Goebeler, V.; Kappelhoff, R.; Butler, G.S. Macrophage matrix metalloproteinase-12 dampens inflammation and neutrophil influx in arthritis. *Cell Rep.* **2014**, *9*, 618–632. [[CrossRef](#)]
73. Jefferson, T.; Čaušević, M.; auf dem Keller, U.; Schilling, O.; Isbert, S.; Geyer, R.; Maier, W.; Tschickardt, S.; Jumpertz, T.; Weggen, S.; et al. Metalloprotease Meprin β Generates Nontoxic N-terminal Amyloid Precursor Protein Fragments in Vivo. *J. Biol. Chem.* **2011**, *286*, 27741–27750. [[CrossRef](#)]
74. Marshall, N.C.; Thejoe, M.; Klein, T.; Serapio-Palacios, A.; Santos, A.S.; von Krosigk, N.; Kizhakkedathu, J.N.; Stoyinov, N.; Foster, L.J.; Overall, C.M. Master Sculptor at Work: Enteropathogenic Escherichia coli Infection Uniquely Modifies Mitochondrial Proteolysis during Its Control of Human Cell Death. *Msystems* **2020**, *5*, e00283-20. [[CrossRef](#)] [[PubMed](#)]
75. Jagdeo, J.M.; Dufour, A.; Fung, G.; Luo, H.; Kleifeld, O.; Overall, C.M.; Jan, E. Heterogeneous Nuclear Ribonucleoprotein M Facilitates Enterovirus Infection. *J. Virol.* **2015**, *89*, 7064–7078. [[CrossRef](#)] [[PubMed](#)]
76. Griswold, A.R.; Cifani, P.; Rao, S.D.; Axelrod, A.J.; Miele, M.M.; Hendrickson, R.C.; Kentsis, A.; Bachovchin, D.A. A chemical strategy for protease substrate profiling. *Cell Chem. Biol.* **2019**, *26*, 901–907.e6. [[CrossRef](#)] [[PubMed](#)]
77. Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G.R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J. Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: Identification of more than 800 *Escherichia coli* proteins. *Mol. Cell. Proteom.* **2002**, *1*, 896–903. [[CrossRef](#)]
78. Van Damme, P.; Van Damme, J.; Demol, H.; Staes, A.; Vandekerckhove, J.; Gevaert, K. A review of COFRADIC techniques targeting protein N-terminal acetylation. *BMC Proc.* **2009**, *3* (Suppl. 6), S6. [[CrossRef](#)] [[PubMed](#)]
79. Perrar, A.; Dissmeyer, N.; Huesgen, P.F. New beginnings and new ends: Methods for large-scale characterization of protein termini and their use in plant biology. *J. Exp. Bot.* **2019**, *70*, 2021–2038. [[CrossRef](#)]
80. Venne, A.S.; Solari, F.A.; Faden, F.; Paretti, T.; Dissmeyer, N.; Zahedi, R.P. An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in *Arabidopsis thaliana*. *Proteomics* **2015**, *15*, 2458–2469. [[CrossRef](#)]
81. Venne, A.S.; Vögtle, F.N.; Meisinger, C.; Sickmann, A.; Zahedi, R.P. Novel Highly Sensitive, Specific, and Straightforward Strategy for Comprehensive N-Terminal Proteomics Reveals Unknown Substrates of the Mitochondrial Peptidase Icp55. *J. Proteome Res.* **2013**, *12*, 3823–3830. [[CrossRef](#)]
82. Venne, A.S.; Zahedi, R.P. The potential of fractional diagonal chromatography strategies for the enrichment of post-translational modifications. *EuPA Open Proteom.* **2014**, *4*, 165–170. [[CrossRef](#)]
83. Jiang, Z.H.; Wu, J.Y. Alternative splicing and programmed cell death. *Proc. Soc. Exp. Biol. Med.* **1999**, *220*, 64–72.
84. Kleifeld, O.; Doucet, A.; Prudova, A.; auf dem Keller, U.; Gioia, M.; Kizhakkedathu, J.N.; Overall, C.M. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* **2011**, *6*, 1578–1611. [[CrossRef](#)]
85. Kleifeld, O.; Doucet, A.; Auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R.K.; Starr, A.E.; Foster, L.J.; Kizhakkedathu, J.N.; Overall, C.M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **2010**, *28*, 281–288. [[CrossRef](#)]
86. Mommen, G.P.M.; van de Waterbeemd, B.; Meiring, H.D.; Kersten, G.; Heck, A.J.R.; de Jong, A.P.J.M. Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO₂-based depletion. *Mol. Cell. Proteom.* **2012**, *11*, 832–842. [[CrossRef](#)]
87. Schlage, P.; Kockmann, T.; Kizhakkedathu, J.N.; Auf dem Keller, U. Monitoring matrix metalloproteinase activity at the epidermal-dermal interface by SILAC-iTRAQ-TAILS. *Proteomics* **2015**, *15*, 2491–2502. [[CrossRef](#)]
88. Mahrus, S.; Trinidad, J.C.; Barkan, D.T.; Sali, A.; Burlingame, A.L.; Wells, J.A. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **2008**, *134*, 866–876. [[CrossRef](#)] [[PubMed](#)]
89. Julien, O.; Zhuang, M.; Wiita, A.P.; O'Donoghue, A.J.; Knudsen, G.M.; Craik, C.S.; Wells, J.A. Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2001–E2010. [[CrossRef](#)] [[PubMed](#)]
90. Agard, N.J.; Maltby, D.; Wells, J.A. Inflammatory stimuli regulate caspase substrate profiles. *Mol. Cell. Proteom.* **2010**, *9*, 880–893. [[CrossRef](#)]
91. Mulvihill, E.E.; Drucker, D.J. Pharmacology, physiology, and mechanisms of action of dipeptidyl peptidase-4 inhibitors. *Endocr. Rev.* **2014**, *35*, 992–1019. [[CrossRef](#)]
92. Shao, W.; Yeretssian, G.; Doiron, K.; Hussain, S.N.; Saleh, M. The caspase-1 digestome identifies the glycolysis pathway as a target during infection and septic shock. *J. Biol. Chem.* **2007**, *282*, 36321–36329. [[CrossRef](#)]
93. Dix, M.M.; Simon, G.M.; Cravatt, B.F. Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* **2008**, *134*, 679–691. [[CrossRef](#)] [[PubMed](#)]