



Degradomics for large-scale mechanistic insights on proteases and proteolysis in human health

Daniel R. Martin,

Sumit Bhutada,

Suneel S. Apte

Department of Biomedical Engineering, Cleveland Clinic Research, Cleveland Clinic, Cleveland, OH, USA

Abstract

Tissue breakdown, especially extracellular matrix or secretome breakdown is a significant aspect of physiological remodeling and disease processes in solid organs, with profound structural and regulatory impact. Tissue and circulating proteins undergo breakdown through a chemical process mediated by proteases, that is, hydrolysis of peptide bonds. Proteolysis has immense biological impact because it is irreversible, results in protein inactivation or activation, and can generate fragments with new functions to expand the functional genome. The traditional focus on a few proteases of interest against candidate substrates provided limited insights into the proteolytic landscape of human diseases. In contrast, innovations in protein terminomics workflows, tandem mass spectrometry and data handling now routinely permit identification of proteolytic events and proteases on the proteome scale, the degradome. The unbiased, *de novo* elucidation of disease degradomes, termed forward degradomics, has increased the number of known *in vivo* proteolytic events, despite only limited application to human disease to date. In reverse degradomics, activities of proteases are elucidated individually, but also at the proteome scale by digesting protein libraries sourced from tissues and cell secretomes, or by comparing the degradomes of protease-deficient/overexpressing and parental cells. Integration of forward and reverse degradomes precisely defines protease primary mechanisms in disease. Cross-disease degradome analysis can define disease-relevant, protease-specific biomarkers, identify proteases as appropriate therapeutic targets, and predict cross-organ impact of protease inhibitors. A systematic effort to map disease degradomes, that is, a prospective human degradome project would generate a comprehensive proteolysis knowledgebase for diagnostics and therapeutics.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Correspondence: S. S. Apte, Department of Biomedical Engineering-ND20, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195, USA., aptes@ccf.org.

Daniel R. Martin and Sumit Bhutada contributed equally to this article.

Author contribution

All authors contributed to conceptualization, writing, editing, and final review of this manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

mass spectrometry; protease; proteolysis; proteomics; terminomics

Introduction

Decoding protein post-translational modifications, especially of extracellular matrix (ECM) components, remains a pressing challenge for the postgenome era. One such modification is proteolysis, which being irreversible, places an indelible mark on ECM by generating proteoforms with new (neo) N- and C-termini. Proteolysis can activate gene products, inactivate them, release bound ligands, modify adhesive sites, and generate bioactive fragments. It is widespread and underestimated; one study found that nearly 50% of proteins in skin had stable cleavage products [1], reflecting the ongoing remodeling of both healthy and disease tissues. ECM proteolysis has enormous biological significance in disease, not only in those conditions where tissue destruction is a principal feature, for example, osteoarthritis, osteoporosis, invasive and metastatic cancers, pulmonary emphysema, and aortic aneurysms, but also in other common disorders, including autoimmune/inflammatory diseases, myocardial infarction, atherosclerosis, and fibrotic conditions affecting a variety of organs. Since cells and their ECM comprise a functional unit, with ECM acting as a repository of information, proteolysis can potentially edit and/or release this information, with powerful regulatory consequences. Despite this impact, the role of proteases in ECM dynamics is poorly resolved in human disease, with few studies conducted at the proteome-wide level. Traditionally, protease research relied on digestion of selected substrates or synthetic peptide libraries by hand-picked proteases resulting in a limited view of disease-relevant cleavage events. Traditional hypothesis-driven research typically also focused on the trendiest proteases and substrates. For example, the decade between 1995 and 2005 saw a disproportionate focus on MT1-MMP in the context of cancer and the proteolytic activity termed aggrecanase (ADAMTS4 and ADAMTS5) in osteoarthritis. During this period and subsequently, much of the protease and ECM space received far less attention.

Association of proteases with diseases has often relied on mRNA detection by RT-qPCR, cell lineage-specific gene expression described by transcriptomics and protein localization by immunostaining. None of these are convincing surrogates for protease activity, since proteases are typically synthesized as zymogens which require activation and most have ambient inhibitors. In the context of the current emphasis on transcriptomics in biomedical research, the cellular origin of proteases may also be misleading, since secreted proteases act noncell-autonomously; an illustrative example is that of clotting factors, most of which are proteases synthesized in the liver that circulate in blood and are active at a distant injury site. The only true measure of ECM proteolysis is a cleaved peptide bond or protein fragmentation observed by gel electrophoresis, but the landscape of proteolytic cleavages is still rather barren, especially in human disease. Which sites in ECM proteins are cleaved? Which proteases are the effectors? How active are these proteases in healthy vs disease tissues? Are specific proteolytic cleavages unique to disease states and what can they tell us about disease status, prognosis and responsiveness to treatment? Since few ECM-modifying

proteases have stringent consensus cleavage site sequences, uncovering protease activity has largely not been amenable to *in silico* prediction and requires experimental evidence.

The advent of N-terminomics, especially the widespread application of Terminal Amine Isotopic Labeling of Substrates (TAILS) [2], demonstrated that despite inherent technologic limitations and ongoing challenges outlined later in this review, it was possible to develop a proteome-scale ‘degradome’ using tissue, cells, or body fluids as starting material and greatly expand our knowledge of proteolysis. This primary annotation of naturally occurring proteolysis in complex proteomes such as tissues and body fluids, using a variety of methods outlined in Table 1 is referred to as ‘forward degradomics’. ‘Reverse degradomics’ uses quantitative strategies applying the same methods for protease-selective, unbiased substrate discovery using mass spectrometry after digestion of a proteome containing relevant substrates with a protease of interest. Higher abundance of peptides with neo-termini compared to those generated in digests with an inactive form of the protease or with an added inhibitor identifies sites of cleavage by the protease. Reverse degradomics can also be undertaken by quantitative comparison of proteomes of cell cultures and tissue from gene knockouts or with engineered over-expression against corresponding wild-type samples (Fig. 1). The purpose of this review is to provide an overview of key methodological concepts and experimental design considerations and illustrate the application of degradomics directly to human health challenges using representative examples. Degradomics using experimental animal models of human disease is not covered in this review unless it serves to illustrate specific concepts.

DEGRADOMICS strategies employing TERMINOMICS with protein labeling

Terminomics uses specialized proteomics approaches for defining protein termini on the proteome scale. The simple principle underlying terminomics is that every intact protein and proteoglycan has one N and C terminus. Each proteolytic cleavage generates an additional N- and C terminus. Degradomics is therefore based on N- and C-terminomics which identifies protein termini and based on the location of the observed terminally labeled peptides with respect to the protein translation start, determines whether they are proteolytic in origin. When signal peptide and propeptide processing generates termini, these, along with the start codon and last amino acid, are referred to as natural termini. Protein termini other than these, typically located further downstream in the protein sequence, are referred to as neo-termini and are presumably proteolytic, since unannotated alternative start sites are rare. Thus, positional peptide annotation is an important step in degradomics (Fig. 1). N-termini are more chemically reactive, and since, at least in theory, both N- and C-terminomics should identify the same cleaved peptide bonds, N-terminomics is more widely used. A number of N- and C-terminomics degradomics strategies are listed in Table 1.

Terminomics relies on liquid chromatography, tandem mass spectrometry (LC–MS/MS), which requires explanation of terms commonly used in this technology. The typical application of LC–MS/MS to proteins is as ‘*shotgun proteomics*’, where the proteome of interest is digested, typically with trypsin, to generate peptides, which are injected into the mass spectrometer, where mass/charge ratio and fragmentation pattern provide the

peptide sequence [3]. Trypsin cleaves proteins predictably after lysine and arginine residues, providing predictable peptides between 7 and 50 amino acids. Methods that utilize peptides to identify proteins in this way are also referred to as ‘*bottom-up*’ proteomics [3,4]. Until recently, such analysis has commonly used *data-dependent acquisition (DDA)*, where a limited number of the most abundant ions detected in the MS1 phase (the data) are selected for a second round of fragmentation (MS2) [3,4]. While DDA is robust, low-abundance ions are frequently underrepresented, because the mass spectrometer favors the most abundant peptides. This challenge is particularly evident in analysis of complex mixtures of molecular components with a high dynamic range of abundance such as tissues and body fluids and is referred to as *high proteome complexity*. It can be mitigated by tissue extraction methods that favor isolation of specific component groups (prefractionation, for example, proteoglycan isolation by ion-exchange chromatography), by depletion of the most abundant proteins, or offline fractionation of peptides prior to LC–MS/MS (Fig. 1). Recently, use of *data-independent acquisition (DIA)* is expanding [5,6]. In DIA, the mass spectrometer sequences all peptides detected in overlapping narrow mass windows, regardless of their relative abundance. This is advantageous over DDA methods, but it presents a new challenge in distinguishing individual peptide sequence information from the spectra of tens to potentially hundreds of different peptides. This computational challenge has been addressed over the past decade with a focus on identifying peptides that specifically - arise from trypsin digestion, which is typically used in bottom-up proteomics [7]. Thus, proteolytic peptides, which can have only one or no tryptic termini, can still be missed because of their lower abundance and computational suboptimization. Therefore, approaches for annotating and enriching proteolytic termini/peptides were necessary and provide the basis of degradomics.

Experimental design considerations

In Fig. 1, we provide an outline of experimental design considerations, which in our experience, heavily influence the outcome. In terminomics with labeling, whether used for forward or reverse degradomics, there are a number of decision points to be considered (Fig. 1). A critical step after protein precipitation in TAILS and HUNTER, for example, is experimental labeling of termini at the level of proteins, before they are reduced to peptides using trypsin, otherwise the existing termini would be invisible against the background of new tryptic termini. Labelling not only adds a mass tag, but also blocks the N-termini, whereas tryptic peptides retain reactivity. Proteolytic peptides are normally of lower abundance in the mixture of peptides generated by trypsin or other ‘working’ proteases; therefore, an enrichment step is usually needed, constituting a *de facto* fractionation step. The methods used for enrichment fall into two categories: positive or negative selection of the labeled/proteolytic peptides. In some conditions, neo-termini are identifiable in shotgun MS datasets by semi-tryptic data analysis, which we discuss separately as an unlabeled degradomics approach. Other review articles that focus on terminomics technologies [8,9] and the original articles cited in Table 1, provide greater detail of each workflow. Terminomics workflows for tissue biopsies, or of body fluids as ‘liquid biopsies’ have also been the subject of excellent reviews [10–12].

The time from tissue harvest to freezing should be minimal and broad-spectrum protease inhibitors should be added when tissue is thawed for forward degradomics. Prior to reverse degradomics, inhibitors for the protease class being tested should be omitted, but inclusion of inhibitors of other classes can be added to minimize 'indirect' cleavages. For reverse degradomics, library selection is a crucial factor in determining which substrates are discovered and how readily. A key concept in defining the repertoire of cleaved secreted proteins is that the cell source is less relevant than its secretome/matrisome, for example, any cell line with an overlapping matrisome repertoire could be used for determining activity of a protease expressed by vascular endothelium or smooth muscle. Because of protease redundancy, tissues or cells with knockout of only one gene of a homologous pair may have sufficient background proteolysis from the intact homolog to conceal authentic proteolytic events. In such instances, double knockouts should be used [1,13]. Tissues such as cartilage which has low cellularity, have a very high percentage of ECM and are well-suited for uncovered ECM proteolysis [14]. On the other hand, highly cellular tissues such as solid tumors have a large dynamic range arising from high cellular protein abundance. For such tissues, decellularization, extensive peptide fractionation, or other measures to compress the dynamic range, should be considered.

Terminomics traditionally required a minimum of 150–300 µg of total protein, two orders of magnitude greater than that required for shotgun proteomes, because only a fraction of the peptides are proteolytic in origin and 90% of the peptides, typically tryptic peptides, are discarded during the enrichment step. However, some workflows such as HUNTER (Table 1) and analysis on the newest mass spectrometers can provide high quality data with much less protein. Relevant to sample proteome complexity, body fluids typically contain several serum-derived proteins such as albumin at high concentration, leading some to employ depletion of the most abundant proteins in order to compress the dynamic range. Offline peptide fractionation by chromatographic methods and separation within the mass spectrometer using high-field asymmetric waveform ion mobility spectrometry (FAIMS) have also been shown to increase the yield of terminome peptides in complementary ways [15]. High dynamic range appears to be less of a hindrance to detection of low-abundance peptides on the newest mass spectrometers with improved speed and depth of analysis.

Proteome isolation for forward degradomics (on ice and with protease inhibitors) can be done with or without fractionating tissue proteins or enrichment of the matrisome, for example, detergent extraction of tissue to minimize cellular proteins, followed by salt extraction (for relatively soluble ECM proteins) and guanidine extraction of insoluble ECM pellets. In addition to serial extraction of ECM and associated molecules, selective enrichment of N-glycosylated proteins (with lectins) and proteoglycans (using anion exchange) could be considered to overcome the large dynamic range in tissues and body fluids. In methods such as TAILS and HUNTER, N-terminal labeling using formaldehyde (reductive dimethylation) provides up to three isotopes for labeling samples individually for duplexing [16]; for experiments that involve quantitative comparison of a larger number of samples, they can be multiplexed for LC–MS/MS analysis after labeling individually with isobaric TMT tags. In TAILS, tryptic peptides are removed by negative selection with hyperbranched polyglycerol aldehydes (HPG-ALD) and the residual N-terminome is used for LC–MS/MS. A portion of the pre-HPG-ALD sample (pre-TAILS) is used for full

proteome mapping and usually yields additional peptides with neo-termini. Labeling and pull-down efficiency are easily determined from the TAILS and pre-TAILS samples for experimental optimization of troubleshooting.

Shedding of the ectodomains of transmembrane proteins is an important area of protease research because of its broad biologic impact and presents a special challenge for degradomics. Transmembrane topology must be considered in attempting to identify the proteolytic products, for example, type I transmembrane proteins have an extracellular N terminus and single ectodomain cleavages would be undetectable by N-terminomics of culture medium or body fluids, whereas the converse is true for type II proteins and the orientation of some multi-pass membrane proteins is not yet known. Methods have been developed to define N-termini displayed on the cell surface [17] or identify the shed proteins, although not their termini [18]. Tsumagari et al used both N- and C-terminal peptide enrichment to identify a number of shed proteins, including several whose shedding was responsive to a metalloprotease inhibitor [19]. Thus, an important consideration in the context of protein shedding experimental design is whether to analyze cell lysate or medium or both using terminomics.

Degradomics data processing

Following data acquisition, mass spectra are deconvolved by dedicated proteomics software (see Table 2) to rigorously identify and quantify peptides. For DDA experiments, options include commercial platforms such as Proteome Discoverer™, which is native for Orbitrap instruments and SPECTROMINE™, Peaks, and open-source tools like MAXQUANT and FRAGPIPE [20]. For data-independent acquisition (DIA) analysis, SPECTRONAUT® is widely used, as are available open-source alternatives including FRAGPIPE [21] and DIA-NN [22]. The latest protein database appropriate for the species of interest is downloaded from UniProtKB/Swiss-Prot for up-to-date searches when the proteome species is known (see section on host–pathogen interactions). Since trypsin is widely used as the working protease, searches are usually set to semi-tryptic, which additionally identifies semi-tryptic C-termini. Carbamidomethylation on cysteines is usually specified as a fixed modification (given standard protein disulfide reduction and alkylation), with variable modifications set to methionine oxidation, N-terminal acetylation, and glutamate/glutamine to pyroglutamate conversion on the peptide termini and isobaric tags or dimethyl on peptide N-termini and lysine. Trypsin missed cleavages are typically set to at least 2, since the blocked lysines resulting in N-terminal labeling methods such as TAILS and HUNTER will result in longer peptides than is usual after trypsin digestion.

De novo sequencing is a specialized mass spectrometry data analysis in which peptide sequence is deduced without a reference database [23]. It is valuable in several contexts, such as expanding the yield of identified peptides when an organism's proteome is unavailable or incompletely annotated, for identification of non-tryptic peptides, such as immunopeptides, or analysis of mixed proteomes, for example, infected tissues. New sequencing algorithms and deep learning packages for *de novo* sequencing are emerging [24,25] that are likely to lead to increased adoption in terminomics, particularly when used in species other than humans and mice.

Annotating peptides on proteins to identify cleavage sites

Downstream tools available for peptide annotation (see Table 2) include MANTI [26] used for MaxQuant output, TermineR [27] for FRAGPIPE and CLIPPER [28] for Trans Proteomics Pipeline [29]. CLIPPER 2.0 [30] uses Proteome Discoverer and Spectronaut data and can be customized for any software that provides information for normalized peptide intensities, accession number, and modifications. These tools can be used for peptide annotation, statistical analysis, and limited data visualization. Once annotation is complete, Perseus [31] can also be used to perform statistics and data visualization. Both N/C-terminal labeling efficiency and the proteolytic peptide enrichment factor should be estimated whenever possible for method optimization. After initial data analysis in reverse degradomics identifies neo-termini, it is necessary to filter out those that arise in proteins that do not occupy the same cellular compartment as the protease of interest. Cytosolic and nuclear proteins, for example, would not typically be relevant substrates of a secreted protease, although intracellular ‘moonlighting’ by secreted matrix metalloproteinases has been previously identified [32].

Public resources for terminomics

UNIPROTKB is the major resource for protein and peptide annotation, specifying signal peptide, transit peptide, and propeptide sequences and protein isoforms arising from alternative translation start sites or splicing events, which can create distinct termini for a given protein. UNIPROTKB provides access to structural information regarding protein domains, post-translational modifications and functional regions which are useful as a reference map, lists known genetic variants and summarizes known functions. MEROPS and TopFIND 4.1 are curated databases that compile experimentally validated cleavage sites allowing forward degradome annotation with known cleavage sites. MEROPS [33] is a peptidase and inhibitor database that organizes proteases into families and clans based on sequence homology, identifying evolutionary relationships and functional similarities. It provides extensive descriptions for each protease, including known substrates, specific inhibitors, and relevant scientific literature. It is potentially useful to identify candidate proteases responsible for observed cleavage events based on substrate specificity patterns and evolutionary relationships. TopFIND 4.1 is a major knowledgebase for N-terminomics research containing protein N-termini from experimental observations, predicted cleavage sites, and alternative translation start sites and splicing events [34]. The database retrieves information from various resources such as experimental N-terminomics, UNIPROT, MEROPS, and community submissions, then categorizes termini as either ‘observed’ (from direct experimental evidence) or ‘inferred’ (from predicted protease cleavage sites or other computational analyses). TopFIND 4.1 provides additional tools such as PATHFINDER which can explore complex, indirect proteolytic pathways.

Database of Identified Cleavage Sites Endemic to Diseases (DICED) [35] is a recently introduced resource for terminomics that provides open access to N- and C-terminal peptide data from diseased tissues and reverse degradomes obtained by the authors. The database can be searched using UNIPROT accession number, protein name, gene symbol, or up to 100 peptide sequences to find experimentally identified cleavage sites. DICED is intended

to facilitate identification of disease-associated protease activity patterns and comparisons of proteolytic events across diseases, tissues, body fluids, and species.

Label-free TERMINOMICS and related approaches

Most LC–MS/MS-based shotgun proteomics analyses use trypsin as the working protease. Since trypsin specifically cleaves after Arg or Lys, peptides that lack predicted ‘tryptic’ termini (at either or both of the N or C terminus) are very likely to be proteolytic in origin. Bell and Overall found that 36% of the N-termini identified by TAILS could also be identified by semi-tryptic peptide searches [36]. Since many proteases cleave after Arg or Lys, semi-tryptic peptide analysis alone will miss such cleavages. Nevertheless, semi-tryptic searches were productively used for targeted analysis to identify fibronectin as a substrate of the proteases ADAMTS9 and ADAMTS16 [33,34]. Ryden et al. analyzed shotgun proteomes of synovial fluid samples taken from patients with knee osteoarthritis and identified numerous nontryptic and semi-tryptic peptides indicative of several known proteolytic activities [32]. They developed the proteasy R package to retrieve experimentally annotated proteases and respective cleavage site data from the MEROPS database for facile cross-analysis with shotgun proteomes. Such terminomics data mining could be used more widely to take advantage of numerous archival shotgun proteomics datasets deposited in ProteomeXchange.

A quantitative LC–MS/MS-based semi-tryptic peptide identification approach, termed the *z*-score method [37], was developed to attribute proteolytic events to specific proteases without prior labeling and enrichment. This label-free approach uses background (tryptic) peptide ratios and abundance to determine significance levels of semi-tryptic peptides. As in other semi-tryptic approaches, trypsin-like cleavages are obscured, potentially providing a skewed view of protease specificity. Tryptic peptides with reduced abundance can support the presence of a cleavage site in that peptide. The experimental set-up requires parallel digestion of a substrate or small library of substrates with an active and inactive protease. The inactive protease is ideally genetically modified but can also be heat denatured or chemically inactivated. Ideally, a proteolytic event should produce 3 unique peptide sequences, that is, 2 peptides arising from the N- and C-terminal side of the cleavage with higher abundance in the active protease digest and a tryptic peptide that spans the cleavage site and has lower abundance in the active protease digest. Therefore, both tryptic and semi-tryptic peptides undergo statistical analysis to identify proteolytic events. A tryptic peptide with higher abundance in the inactive protease digest supports the occurrence of proteolytic degradation within its sequence but does not provide the precise cleavage site. To disclose trypsin-like activities, an alternative working protease such as Glu-C can be employed with appropriate semi-Glu-C data searches. Table 1 lists several terminomics approaches that do not utilize N- or C-terminal labeling but are more complex than semi-tryptic site analysis from shotgun proteome datasets since they employ one or more additional processing/enrichment steps.

Peptidomics

Label-free detection and sequencing of small peptides can complement degradomics methods that use bottom-up proteomics since initial size fractionation in the peptidomics

sample preparation workflow, such as by molecular weight cut-off filters, excludes large proteins and large fragments. Selecting a pool of small peptides in this way reduces sample dynamic range and increases the likelihood of peptide detection; however, peptide identification requires database searches with no protease specified and may require *de novo* sequencing. Furthermore, peptidomics is limited to small fragments of proteins and may rely on endopeptidase cleavages occurring sufficiently near each other. It may also be necessary to distinguish such products of regulated protein processing from protein degradation products. In a recent study, saliva from patients with oral squamous cell cancer was used for a coupled peptidome-proteome analysis [38], which identified numerous proteolytic peptides and demonstrated a reduction in protease inhibitors in saliva from patients with lymph node-positive cancer. Plasma and lymph peptidomics have been used to identify circulating endogenous or self-antigens relevant to the major histocompatibility complex and hence immune tolerance and autoimmune conditions; many of these may be proteolytic in origin [39]. Clement et al [40] found that in addition to the endosomal processing pathway that generated cathepsin-cleaved peptides, the peptidome eluted from HLA-DR1 receptor on dendritic cells contained numerous small peptides that were attributed to diverse metalloproteases such as MMPs and ADAMs. Peptidomics of a size-selected protein fraction (< 5 kDa) identified numerous peptides in human pancreatic cyst fluid, the majority arising by cleavage of plasma proteins [41]. Since pancreatic cysts contain several proteases, the peptides report their activity, which is potentially useful for diagnostic purposes. Alignment of semi-tryptic peptides and peptidomes to the human plasma peptide atlas, for example, would be useful for prioritization of cleaved peptides likely to be detected in the circulation [42].

Protein N-terminal sequencing

Despite advances in mass spectrometry techniques, N-terminal protein sequencing using Edman degradation [36] has a necessary role in investigating ECM proteolysis. This is because LC-MS/MS detection of peptides with extensive PTMs or with Arg residues adjacent to or distant from the cleaved site (resulting in peptides outside the set instrument mass range) may be challenging. N-terminal protein sequencing was used to uncover the canonical cleavage site of ADAMTS proteases in the interglobular region of aggrecan, and to identify additional cleavages in the heavily modified central regions of aggrecan and versican [43,44]. Although these sites were mapped before the advent of mass spectrometry terminomics, their discovery would be unlikely using mass spectrometry-based terminomics as discussed later in this review. Protein fragments used for Edman degradation must be pure and in sufficient quantity and typically need to be bound to polyvinylidene fluoride membranes after electrophoretic separation. The N terminus is determined by stepwise amino acid labeling and cleavage, followed by chromatography of the eliminated residues to determine their molecular mass.

Investigation of human health and diseases using DEGRADOMICS

Pathogen-host interactions

Bacterial and helminth (parasitic worm) pathogens produce proteases that can act as virulence factors by modulating host immune responses and destroying host tissues. Peptides

generated by pathogen proteases are likely detected during degradomics of infected tissues or healthy surface tissues and their microbiomes, but the responsible proteases can be difficult to identify, in the absence of reverse degradomics analysis of pathogen proteases. This is especially true in mixed pathogen colonization, such as of open wounds or in conditions which have similar clinical presentations despite infections with different pathogens, such as infective endocarditis. Infective endocarditis is a dangerous condition that results from systemic infection with a variety of pathogens and is characterized by formation of wart-like vegetations on the surface of heart valves, which are locally destructive to valve and surrounding tissue, interfere with valve function and may embolize to cause infective thrombi at distant locations. Among responsible pathogens, the most common is *Staphylococcus aureus*, although Gram-negative organisms and fungi are also implicated. The vegetations represent the outcome of a host–pathogen interaction on the endothelial surface of heart valves. Martin et al determined the degradome and proteome of vegetations and found a large number of proteolytic peptides [45], but attribution to proteases was not undertaken. Reverse degradomics of the staphylococcal V8 protease (commonly used as endoproteinase GluC for proteome digestion in proteomics workflows) against the neutrophil proteome was achieved by Mustor et al [46], using TAGS-CR terminomics (Table 1), identifying 350 new substrates, including proteins with a critical role in neutrophil defense. Using aminobiotinylation and TAILS against the plasma proteome, 90 substrates directly relevant to *S. aureus* pathogenesis were identified including proteins of the clotting cascade, protease inhibitors and the complement system [47], consistent with proteome analysis of vegetations, which showed an abundance of cleaved circulating and neutrophil-derived proteins [45].

Microbial proteases may be involved in homeostatic proteolysis occurring at interfaces having a substantial microbiome population such as colonic, bronchial, and vaginal mucosa [48]. Thus, proteolytic activity from healthy microbiomes could very well contribute to good health and this could be an area of future research. For example, meprin β -mediated MUC2 cleavage regulates mucus detachment in the intestine and prevents bacterial overgrowth [49]. *Porphyromonas gingivalis* infection of neutrophils or processing of the neutrophil proteome by its protease RgpB, a gingipain, led to cleavage of several azurophilic granule proteins whose neoepitopes, translocated to the surface, activated macrophage internalization of live neutrophils. Macrophages with entrapped neutrophils had subdued proresolution pathways, hindering rapid resolution of inflammation [50].

Degradomics was used to elucidate the role of the 3CL main protease (3 CL M^{Pro}) produced by the SARS-CoV-2 virus responsible for the recent COVID-19 pandemic. Pablos et al. used TAILS for degradomics to identify cleaved human proteins in human embryonic kidney cells and IFN- α treated human lung epithelial (BEAS-2B) cells infected with SARS-CoV-2 to identify 102 cleavages in 101 proteins [51]. This analysis identified an impact on the Hippo signaling pathway, thereby affecting cell morphology, mechano-transduction, tissue growth, and regeneration. One of the manifestations of COVID-19 is a hypercoagulable state. Pagotto et al. showed that M^{Pro} could directly activate coagulation factors VII and XII, each a protease [52]. Degradomics also showed that 3 CL M^{Pro} had an extended secondary specificity for cleavage at R↓X that was distinct from its known preference for LQ↓X sites [51,52].

Although not a host–pathogen interaction, snakebites by Viperidae species, which inject hemorrhagic venoms, are an important public health challenge and significant cause of mortality in many parts of the world. The envenomation exposes the human proteome to proteolysis at the bite site and in distant organs such as the kidney [53]. Hemorrhagic snake venoms contain an array of proteases, some of which are metalloproteinases collectively grouped as reprotolysins, whose catalytic domains resemble those of human ADAMs and ADAMTS proteases. Administration of venom from the highly venomous south American pit viper *Bothrops jararaca* identified thousands of proteolytic cleavages in the kidneys, with a large number attributed to activation of host proteases [53]. An evolving profile of proteolytically cleaved sequences with time after the envenomation suggested activation of a number of host proteases or destruction of their inhibitors. A high frequency of cleavages observed after asparagine residues specifically suggested activation of endogenous legumain [53]. In another study, the activity of the protease HF3 from this snake species was defined against human plasma after depletion of abundant components and peptidomics was applied to identify the proteolytic products [54]. Such analyses are valuable because they could provide biomarkers for severity of envenomation and could assist in development and testing of antivenoms, including *de novo* designed proteins [55].

Human tissue and fluid degradomes

Animal models of human diseases have been subjected to terminomics, such as a rat ischemia–reperfusion injury model [56], a genetic pancreatic tumor model [57] and an inflammatory colitis model [58], but because such models generally do not faithfully replicate human pathogenesis, this section will focus on direct analysis of human material. Surprisingly, little terminomics analysis has been done in human tumors. Song et al. developed a cleavable 2-pyridinecarboxyaldehyde probe (2PCA-Probe) for N-terminome determination and a new marker identifying N-terminal chemical labeling with the 2PCA-Probe and applied the method to identify 4686 N-terminal peptides in colorectal cancer and adjacent tissues [59].

Lorentzen et al. utilized a label-free N-terminomics approach after digestion of proteins from carotid artery plaques obtained from 21 patients via endarterectomy [60]. The study identified over 4000 N-terminal peptides determined to arise from proteolysis. Analysis of these using TopFIND identified 567 sites that could be attributed to known protease activities, involving meprins, MMPs, chymase, and neutrophil elastase [60]. However, the majority of sites were unattributable to a specific protease, emphasizing the need for acceleration of reverse degradomics.

Human body fluids are well amenable to terminomics analysis despite their high protein dynamic range, which can be mitigated by prefractionation of the sample, depletion of the most abundant proteins such as albumen or fractionation of the peptides after proteome digestion. Sabino et al. identified 9802 neo-termini arising from 2020 proteins in human wound fluid extracted from foam dressings [61]; they reduced proteome complexity by combinatorial peptide ligand libraries (available commercially as Proteomimer) [62]. Wildes and Wells used N-terminal labeling by subtiligase to undertake terminomics of human

blood, identifying 772 unique N-termini in 222 proteins and sampling across six orders of magnitude of abundance [63].

Das and colleagues performed the first degradomics analysis of synovial fluid and demonstrated that mast cell tryptase degraded a key joint component, proteoglycan-4 or lubricin, which serves as an essential lubricant [64]. Bhutada et al used TAILS to investigate proteolysis in one of the most common diseases associated with tissue breakdown, that is, osteoarthritis (OA). In this condition affecting over 7% of the world's population, especially adults over 60 years of age, there is extensive bone and cartilage remodeling and progressive loss of articular cartilage. Analysis was first done using OA cartilage [14] and subsequently, matched cartilage and synovial fluid [65] from patients undergoing knee replacement surgery. Proteolytic destruction of osteoarthritic cartilage was found to occur on an unexpectedly massive scale, with nearly 11 000 proteolytic sites identified in cartilage and over 14 000 in synovial fluid, the majority arising in ECM components (in cartilage) and secreted/plasma proteins (in synovial fluid) [14,65]. Analysis of matched cartilage and synovial fluid samples established that proteolytic cartilage breakdown products were transmitted to synovial fluid with 677 peptides from 153 proteins consistently present in both sample types [14,65]. Together with prior work on synovial fluid by Das et al [64], this work supported synovial fluid analysis as a suitable liquid biopsy of the joint for developing clinically useful biomarkers. A striking feature of the OA degradome was pervasive 'ragging', that is, stepwise deletion of 1–3 amino acids from peptide termini [14,65]. It accounted for approximately 47% of identified cleavages, suggesting widespread exopeptidase activity alongside endopeptidase cleavage events. Ragging likely reflects the prolonged residence time of proteolytic fragments within the joint, where they may be exposed to exopeptidases over the long OA timeline, which may span several years of gradual cartilage destruction and remodeling.

By integrating forward degradomics of late-stage knee OA cartilage with reverse degradomics of individual proteases HtrA1, ADAMTS5, MMP13, and CMA1, Bhutada and colleagues attributed > 1000 specific cleavages in the OA cartilage degradome to distinct proteolytic enzymes [14,65]. HtrA1, identified as the most abundant protease in these samples, generated 323 cleavages from 109 cartilage proteins [14]. Reverse degradomics of ADAMTS5, a primary aggrecanase, revealed 287 cleavage sites with broad activity against multiple substrates, including COMP, PRELP, and various collagens, although its detection of aggrecan peptides was limited by technical challenges in identifying heavily glycosylated regions [65]. MMP13 cleaved collagen II as well as PRELP, biglycan, and COMP. Notably, the mast cell protease CMA1, whose role in OA was previously undefined, generated 201 distinct cleavages predominantly in COMP, aggrecan, PRELP, and fibronectin [65]. Comparing all four protease degradomes revealed very little overlap in the sites they attacked. Furthermore, even when accounting for ragging-related redundancy, the reverse degradomes of HtrA1, ADAMTS5, MMP13, and CMA1 collectively explained only a fraction of the total cleaved peptides observed in OA cartilage and synovial fluid [65]. This finding strongly suggests that cartilage degradation in human knee OA results from a complex, multiple-protease-mediated degradative process with the fragments accumulating over the long disease timeline. Although ADAMTS5 and CMA1 were not detected in late-

stage OA cartilage, the presence of their proteolytic products suggested different protease activities may dominate at distinct phases of disease progression.

HUNTER with a shortened labeling time was used to define the plasma N-terminome in patients with hemolytic uremia syndrome induced by Shiga-toxin E.coli who were undergoing treatment with eculizumab for complement inhibition, comparing plasma samples at baseline and after treatment [66]. This analysis detected 476N-termini from 130 proteins which were differentially abundant, showing that N-terminomics could be effectively applied to study proteolytic impact of drugs using plasma. Next, the plasma N-terminome of patients with systemic lupus erythematosus (SLE) was analyzed using HUNTER optimized with a robot. In most SLE patients, significant changes in proteolytic termini were detected relative to healthy controls, including inter-alpha-trypsin inhibitor heavy chain H1 (ITIH1) a protease inhibitor and PROS1, an immunosuppressive anticoagulant protein [66]. In addition, several N-termini were associated with specific clinical parameters of SLE and the patient trajectory [66]. Reverse degradomics of four proteases active in the human complement system (MASP-1, MASP-3, C1S, C1r) was undertaken by digestion of heat-treated plasma *in vitro*, identifying a number of novel cleavages as well as a significant proportion of cleavages in complement proteins, some of which were detected in SLE plasma [66]. In addition to providing an excellent example of how N-terminomics can be deployed for forward and reverse degradomics, this article made a strong case for expanding degradome research by showing that specific immunoactive fragments that modulate inflammation and tissue injury, such as one from complement C3 characterized therein, can be generated by proteolysis [66].

The limited examples of forward degradomics using human samples will benefit from cross-analysis with reverse degradomics undertaken in human proteome libraries. For example, comparison of cells lacking a protease of interest with the wild-type counterpart has been undertaken for calpain-2 [67], ADAMTS9 [68], and calpain-5 [69]. Anderson et al. defined cleaved sites in human macrophage proteins after chemical inhibition of legumain [70]. TAILS has also been applied for drug testing by reverse degradomics. For example, TAILS of vitreous humor from rabbits and monkeys treated with an HtrA1 inhibitory antibody was used to determine the relevant substrates of HtrA1 and inhibitory efficiency of this antibody [71]. Proteolytic peptides in retinol-binding protein 3 (RBP3), Dickkopf-related protein 3 (DKK3), and clusterin (CLU) were consistently identified in both animal models and were higher in HtrA1-active as compared with HtrA1-inhibited samples.

Technical limitations and current challenges in DEGRADOMICS

Caveats and limitations of terminomics

Maximizing proteome yield from tissues for forward degradomics may require harsh conditions, including extremes of pH, which may hydrolyze peptide bonds and addition of chaotropic agents such as guanidine hydrochloride, which denatures proteins and their complexes. Hence, it may be necessary to accept a residual pellet during tissue extraction to ensure minimal protein disruption. Reverse degradomics usually does not employ similar extraction conditions; therefore, proteolytic cleavages in reverse degradomes that match those in forward degradomes may provide confidence for proteolytic origin of the neo-

peptide. The dynamic range of ECM (as well as of cells, body fluids, and tissues) means that some exceedingly rare neo-peptides may escape detection despite enrichment.

Technical limitations

Many technical limitations of degradomics are related to the inherent drawbacks of shotgun proteomics, which is used in most workflows, since proteolysis with trypsin or other proteases is necessary to digest proteins to peptides for detection. As a result, and in contrast to 'top-down' proteomics and peptidomics, shotgun experiments typically only identify short peptides with neo-termini, but not entire protein fragments (other than peptides with nontryptic N- and C-termini at each end). Peptides which are larger or smaller than the selected mass window, determined by where trypsin cleavage sites occur in the vicinity of fragment ends, are likely to escape detection. In addition, peptides with complex or unknown modifications (e.g., many forms of glycosylation) typically escape detection unless the modification is specifically allowed for. The large aggregating proteoglycan aggrecan, whose breakdown is important in OA pathogenesis, illustrates some of these challenges. In addition to its 3 globular domains, it has a large central glycosylated region with numerous O- and N-glycans and chondroitin sulfate chains. On shotgun analysis of aggrecan (or the closely related proteoglycan versican), coverage of this region is typically sparse and the known cleavages occurring here were uncovered by Edman degradation. For example, the neo N terminus ARGSVIL results from cleavage at a canonical cleavage site (RNITEGE↓ARGSVIL) by ADAMTS proteases known as aggrecanases. When trypsin is used as the working protease, it will cleave two residues downstream of this cleavage, generating an undetectable dipeptide Ala-Arg, explaining why a neo-N-terminal peptide downstream of the cleavage site has not been detected by TAILS analysis of OA cartilage or cartilage digested with ADAMTS5. Even if this trypsin cleavage does not occur (i.e., a 'missed' cleavage), the next Arg/Lys residue is 46 amino acids downstream so that the resulting peptide may be too large to be detected. Upstream of the aggrecanase cleavage site, the peptide with sequence NITEGE is theoretically detectable by LC-MS/MS but was not identified. It is possible that it carries a modification or is further cleaved by another peptidase. In contrast, aggrecan cleavage at the canonical MMP site DIPEN↓FFGVG was detected by identification of a C-terminal peptide with sequence [R].YDAICYTGEDFVDIPEN.[F], whereas the predicted neo-N-terminal peptide, [N].FFGVGGEEDITVQTVTPDMEPLPR was not detected. These examples illustrate how even with the extensive proteome coverage permitted by LC-MS/MS, the scale of proteolysis in any disease is likely to be severely underestimated.

In top-down proteomics, in contrast, digestion with trypsin or another protease is omitted prior to LC-MS/MS, which means that intact protein isoforms and fragments can be observed. Briefly, the protein ion is first stored for mass measurement and then fragmented to determine its sequence. Top-down proteomics presently has limited relevance for ECM because of the extensive and variable PTMs of ECM components and the formation of complexes by most ECM proteins. It is potentially advantageous for the detection of small proteins which may not generate a sufficient number of appropriate tryptic peptides compared to large proteins.

Identifying relevant cleavages and measuring dynamics

Since ECM proteins are modular and the modules frequently contain disulfide bonds, not all proteolytic cleavages would lead to fragmentation. An intra-module cleavage may not fragment the protein if the fragments are held together by disulfide bonds (such a cleavage would result in a 'nick'). Some observed cleavages in reverse degradomics may not be direct substrates of the targeted protease, but indirect, via activation of other proteases, or degradation of inhibitors. For example, it was notable that the well-established site specificity of V8 protease with Glu or Asp at the P1 position was found in only 74% of cleavages observed by V8 digestion of plasma, suggesting that a quarter may have arisen from activation of host proteases [47]. Degradomics does not directly provide information about the kinetics of cleavage, nor can it reliably rank substrate and site preference or how recently cleavage may have occurred, without specific experimental design, such as serial sampling of digests or by use of multiple time points of digestion in reverse degradomics. A time-resolved analysis of the activity of MMP10 reverse degradome was undertaken using multiplexed iTRAQ-TAILS to determine relative proteolytic efficiency and structural accessibility of multiple cleavage events [72]. Stable isotope labeling by amino acids in culture (SILAC) has been used with N-terminomics for a dynamic view of the N-terminome [73], but because labeled amino acids are expensive, it may be more feasible in cell cultures. Tissues from animals fed with specific SILAC labeling regimens could be used for degradomics, but such experiments are unlikely in human disease.

Indirect proteolysis and ragging

Analysis of a reverse degradome identifies cleavage sites occurring downstream of a protease, but not necessarily its direct substrates owing to the possibility of activation of another protease. Follow-up validation using binary digests of purified protease with purified substrate can confirm that the substrate is cleaved directly rather than indirectly. Ragging is the appearance of stepwise loss of 1–2 amino acids from an N or C terminus, resulting in a ladder that becomes evident when viewing neo-peptides in a forward degradome using annotation by sequence number. It is rarely seen in reverse degradomes with endopeptidases, suggesting a role for exopeptidases in further modification of endopeptidase cleavages in tissue/disease proteomes. Activity of endopeptidases acting in close proximity can potentially contribute to the complexity of distinguishing primary cleavages from ragging. One of the benefits of undertaking reverse degradomics is that it identifies the primary endopeptidase cleavages from which ragging may be initiated.

Translation of DEGRADOMES for human disease and future developments

Beyond expanding knowledge of human disease, tissue turnover and protease mechanisms, human degradomics has practical value for diagnosis and treatment. Several ECM-derived matrikines, which are proteolytic fragments with bioactivity are known, for example, endostatin from collagen XVIII [74], endotrophin from the $\alpha 3$ chain of collagen VI [75], and versikine from versican [76]. Peptides with known neo-termini have provided immunogens that can be used to generate antibodies highly specific for cleaved termini but not intact sequences. Such neopeptide antibodies have been generated against a number of termini, proving their utility beyond a doubt, and are versatile reagents that can be used

in immunostaining, western blotting and ELISA. Examples include neoepitope antibodies to cleavages in the proteoglycans aggrecan [44,77], versican [78] and brevican [79], collagens [80,81] and cartilage oligomeric matrix protein (COMP, or thrombospondin 4) [82].

Systematic analysis of human diseases is likely to uncover associations of neo-termini with distinct patterns of proteolysis in human diseases. Clusters of disease-associated neo-termini could be used to generate targeted proteomics assays referred to as multiple reaction monitoring (MRM) or parallel reaction monitoring (PRM). Proteases strongly associated with disease can be targeted by drugs, providing new therapeutic approaches. For now, analyses of cell cultures and mouse models of diseases outnumber those using human tissues and body fluids. However, the necessary elements are in place for systematically mapping the degradomes of different diseases, that is, diverse N-terminomics and C-terminomics workflows, high-resolution mass spectrometers with high-throughput, powerful data analytic packages and routine submission of datasets to ProteomeXchange as a condition of publication.

This ecosystem makes possible consideration of a *human degradome project*, a purpose-driven initiative to map the proteolytic landscapes of a variety of human diseases. It has the potential to provide an abundance of new biomarkers for precision medicine, mechanistic knowledge of protease activities (via reverse degradomics) and identify proteases as relevant drug targets. It would allow determination of overlap and contrasts in disease degradomes, help stratify disease subtypes and elucidate chronic disease timelines, define protease activities across diseases, test inhibitors and help predict inhibitor effects—especially potential side effects at nondisease sites. A knowledgebase on this scale would be transformative for the global medical research community and is a desirable objective for acceleration of diagnostics and medicines. Such an initiative could be implemented for proof of concept and feasibility on a small scale in a single laboratory using a limited variety of diseased tissues or body fluids with public access to data but better implemented by collaborating groups worldwide. It would require consideration and prioritization of the most common disease states (e.g., inflammatory conditions, cancers, degenerative disorders), inclusion of diverse tissue origins (by organ, e.g., breast cancer, colon cancer) and selection of appropriate representative patients by expert clinicians. On the technical side, the project would be facilitated by automated sample handling, access to high-resolution mass spectrometers, a variety of LC–MS/MS data analysis tools and computer hardware and software for storage, retrieval and online access. All these resources exist, but successful implementation will require bringing them together, ideally by collaboration of the interested parties and dedicated funding.

Acknowledgements

The authors acknowledge funding from the US National Institutes of Health (Award ARO85047 to S.A.) and The Marfan Foundation (to D.R.M.). The funding sources had no other involvement in writing of this article.

Data availability statement

This is a literature review and opinion article. Therefore, no data was generated or presented.

Abbreviations

2PCA	2-pyridinecarboxaldehyde
ADAM	A disintegrin and metalloprotease domain
ADAMTS	A disintegrin-like and metalloprotease domain with thrombospondin type 1 motif
CHA-FRADIC	charge-based fractional diagonal chromatography
CHAMP-N	chromatographic amplification of protein N-terminal peptides
CHOPPER	chemical enrichment of protease substrates with purchasable elutable reagents
CHOPS	chemical enrichment of protease substrates
COFRADIC	combined fractional diagonal chromatography
CPB-ChaFRADIC	Carboxypeptidase B-assisted charge-based fractional diagonal chromatography
DDA	data-dependent acquisition
DIA	data-independent acquisition
DIA-NN	data-independent acquisition neural-network
DICED	database of identified cleavage sites endemic to diseases
ECM	extracellular matrix
EDC-	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
FAIMS	field asymmetric ion mobility spectrometry
GAP3	glyceraldehyde-3-phosphate
HPG-ALD	hyperbranched polyglycerol aldehydes
HUNTER	high-efficiency undecanal-based N-termini enrichment
HYTANE	hydrophobic tagging-assisted N-termini enrichment
iTRAQ	isobaric tag for relative and absolute quantitation
LC-MS/MS	liquid chromatography tandem mass spectrometry
MANTI	Maxquant advanced N-termini interpreter
MMP	matrix metalloproteinase
MOLEX	molecular extraction

MRM	multiple reaction monitoring
N-CLAP	N-terminalomics by chemical labeling of the alpha-amine of proteins
NHS	N-hydroxy succinimide
OA	osteoarthritis
PBC	2-PCA- and biotin labeling-based C-terminomics
PICS-2	proteomic identification of cleavage sites with 2PCA reagents
PITC	phenyl isothiocyanate
PRM	parallel reaction monitoring
P-TAG	phosphor-tagging
PTM	post-translation modification
RT-qPCR	reverse transcriptase-quantitative polymerase chain reaction
SARS-CoV2	severe acute respiratory syndrome coronavirus 2
SCX	strong cation exchange
SILAC	stable isotope labeling by amino acids in culture
SLE	systemic lupus erythematosus
TAGS-CR	terminal amine guanidination of substrates-charge reversal
TAILS	terminal amine isotopic labeling of substrates
TFA	trifluoroacetic acid
TMT	tandem mass-tags

References

1. dem auf Keller U, Prudova A, Eckhard U, Fingleton B & Overall CM (2013) Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. *Sci Signal* 6, rs2. [PubMed: 23322905]
2. Kleifeld O, Doucet A, dem auf Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN & Overall CM (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* 28, 281–288. [PubMed: 20208520]
3. Zhang Y, Fonslow BR, Shan B, Baek MC & Yates JR 3rd (2013) Protein analysis by shotgun/ bottom-up proteomics. *Chem Rev* 113, 2343–2394. [PubMed: 23438204]
4. Aebersold R & Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. [PubMed: 27629641]

5. Kitata RB, Yang JC & Chen YJ (2023) Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. *Mass Spectrom Rev* 42, 2324–2348. [PubMed: 35645145]
6. Lou R & Shui W (2024) Acquisition and analysis of DIA-based proteomic data: a comprehensive survey in 2023. *Mol Cell Proteomics* 23, 100712. [PubMed: 38182042]
7. Fröhlich K, Fahrner M, Brombacher E, Seredynska A, Maldacker M, Kreutz C, Schmidt A & Schilling O (2024) Data-independent acquisition: a milestone and Prospect in clinical mass spectrometry-based proteomics. *Mol Cell Proteomics* 23, 100800. [PubMed: 38880244]
8. Haack AM, Overall CM & Auf dem Keller U (2022) Degradomics technologies in matrisome exploration. *Matrix Biol* 114, 1–17. [PubMed: 36280126]
9. Canbay V & Auf dem Keller U (2021) New strategies to identify protease substrates. *Curr Opin Chem Biol* 60, 89–96. [PubMed: 33220627]
10. Anowai A, Chopra S, Mainoli B, Young D & Dufour A (2022) N-Terminomics/TAILS of tissue and liquid biopsies. *Methods Mol Biol* 2456, 85–94. [PubMed: 35612737]
11. Derakhshani A, Bulluss M, Penner R & Dufour A (2024) N-Terminomics/TAILS of human tumor biopsies and cancer cell lines. *Methods Mol Biol* 2747, 19–28. [PubMed: 38038928]
12. Kockmann T, Carte N, Melkko S & Keller U (2016) Identification of protease substrates in complex proteomes by iTRAQ-TAILS on a Thermo Q Exactive instrument. In *Neuromethods* (Grant JE & Li H, eds), pp. 187–207. Springer Science+Business Media, New York.
13. Mead TJ, Bhutata S, Peruzzi N, Adegboye J, Seifert DE, Cahill E, Drinko J, Donnellan E, Guggiliam A, Popovic Z et al. (2025) ADAMTS7, a target in atherosclerosis, cooperates with its homolog ADAMTS12 to protect against myxomatous valve degeneration. *J Mol Cell Cardiol Plus* 11, 100288. [PubMed: 40115634]
14. Bhutata S, Li L, Willard B, Muschler G, Piuze N & Apte SS (2022) Forward and reverse degradomics defines the proteolytic landscape of human knee osteoarthritic cartilage and the role of the serine protease HtrA1. *Osteoarthr Cartil* 30, 1091–1102.
15. Ziegler AR, Anderson BM, Latorre R, McQuade RM, Dufour A, Schmidt BL, Bunnett NW, Scott NE & Edgington-Mitchell LE (2025) N-terminomics profiling of naïve and inflamed murine colon reveals proteolytic signatures of legumain. *J Cell Physiol* 240, e31466. [PubMed: 39392222]
16. Doucet A, Kleifeld O, Kizhakkedathu JN & Overall CM (2011) Identification of proteolytic products and natural protein N-termini by terminal amine isotopic labeling of substrates (TAILS). *Methods Mol Biol* 753, 273–287. [PubMed: 21604129]
17. Lui I, Schaefer K, Kirkemo LL, Zhou J, Perera RM, Leung KK & Wells JA (2025) Hypoxia induces extensive protein and proteolytic remodeling of the cell surface in pancreatic adenocarcinoma (PDAC) cell lines. *J Proteome Res* 24, 2791–2800. [PubMed: 40312771]
18. Tüshaus J, Müller SA, Kataka ES, Zaucha J, Sebastian Monasor L, Su M, Güner G, Jocher G, Tahirovic S, Frishman D et al. (2020) An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *EMBO J* 39, e105693. [PubMed: 32954517]
19. Tsumagari K, Chang CH & Ishihama Y (2021) Exploring the landscape of ectodomain shedding by quantitative protein terminomics. *iScience* 24, 102259. [PubMed: 33796845]
20. Yu F, Deng Y & Nesvizhskii AI (2025) MSFragger-DDA+ enhances peptide identification sensitivity with full isolation window search. *Nat Commun* 16, 3329. [PubMed: 40199897]
21. Yu F, Teo GC, Kong AT, Fröhlich K, Li GX, Demichev V & Nesvizhskii AI (2023) Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. *Nat Commun* 14, 4154. [PubMed: 37438352]
22. Demichev V, Messner CB, Vernardis SI, Lilley KS & Ralser M (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* 17, 41–44. [PubMed: 31768060]
23. König S, Obermann WMJ & Eble JA (2022) The current state-of-the-art identification of unknown proteins using mass spectrometry exemplified on De novo sequencing of a venom protease from *Bothrops moojeni*. *Molecules* 27,
24. Sanders J, Wen B, Rudnick PA, Johnson RS, Wu CC, Riffle M, Oh S, MacCoss MJ & Noble WS (2025) A transformer model for de novo sequencing of data-independent acquisition mass spectrometry data. *Nat Methods* 22, 1447–1453. [PubMed: 40596427]

25. Eloff K, Kalogeropoulos K, Mabona A, Morell O, Catzel R, Rivera-de-Torre E, Berg Jespersen J, Williams W, van Beljouw SPB, Skwark MJ et al. (2025) InstaNovo enables diffusion-powered de novo peptide sequencing in large-scale proteomics experiments. *Nature Machine Intelligence* 7, 565–579.
26. Demir F, Kizhakkedathu JN, Rinschen MM & Huesgen PF (2021) MANTI: automated annotation of protein N-termini for rapid interpretation of N-Terminome data sets. *Anal Chem* 93, 5596–5605. [PubMed: 33729755]
27. Cosenza-Contreras M, Seredynska A, Vogele D, Pinter N, Brombacher E, Cueto RF, Dinh TJ, Bernhard P, Rogg M, Liu J et al. (2024) TermineR: extracting information on endogenous proteolytic processing from shotgun proteomics data. *Proteomics* 24, e2300491. [PubMed: 39126236]
28. dem auf Keller U & Overall CM (2012) CLIPPER: an add-on to the trans-proteomic pipeline for the automated analysis of TAILS N-terminomics data. *Biol Chem* 393, 1477–1483. [PubMed: 23667905]
29. Deutsch EW, Mendoza L, Shteynberg DD, Hoopmann MR, Sun Z, Eng JK & Moritz RL (2023) Trans-proteomic pipeline: robust mass spectrometry-based proteomics data analysis suite. *J Proteome Res* 22, 615–624. [PubMed: 36648445]
30. Kalogeropoulos K, Moldt Haack A, Madzharova E, Di Lorenzo A, Hanna R, Schoof EM & Auf dem Keller U (2024) CLIPPER 2.0: peptide-level annotation and data analysis for positional proteomics. *Mol Cell Proteomics* 23, 100781. [PubMed: 38703894]
31. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M & Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13, 731–740. [PubMed: 27348712]
32. Jobin PG, Butler GS & Overall CM (2017) New intracellular activities of matrix metalloproteinases shine in the moonlight. *Biochim Biophys Acta Mol Cell Res* 1864, 2043–2055. [PubMed: 28526562]
33. Rawlings ND & Bateman A (2021) How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci* 30, 83–92. [PubMed: 32920969]
34. Fortelny N, Yang S, Pavlidis P, Lange PF & Overall CM (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. [PubMed: 25332401]
35. Joshi J, Bhutada S, Martin DR, Guzowski J, Blankenberg D & Apte SS (2025) DICED (database of identified cleavage sites endemic to diseases states): a searchable web Interface for Terminomics/ Degradomics. *Proteomics* 25, e202500007. [PubMed: 40351053]
36. Bell PA & Overall CM (2023) No substrate left behind-Mining of Shotgun Proteomics Datasets Rescues Evidence of proteolysis by SARS-CoV-2 3CL (pro) Main protease. *Int J Mol Sci* 24,
37. Martin DR, Santamaria S, Koch CD, Ahnström J & Apte SS (2021) Identification of novel ADAMTS1, ADAMTS4 and ADAMTS5 cleavage sites in versican using a label-free quantitative proteomics approach. *J Proteome* 249, 104358.
38. Neves LX, Granato DC, Busso-Lopes AF, Carnielli CM, Patroni FMS, De Rossi T, Oliveira AK, Ribeiro ACP, Brandão TB, Rodrigues AN et al. (2021) Peptidomics-driven strategy reveals peptides and predicted proteases associated with Oral cancer prognosis. *Mol Cell Proteomics* 20, 100004. [PubMed: 33578082]
39. Santambrogio L & Rammensee HG (2019) Contribution of the plasma and lymph Degradome and Peptidome to the MHC Ligandome. *Immunogenetics* 71, 203–216. [PubMed: 30343358]
40. Clement CC, Becerra A, Yin L, Zolla V, Huang L, Merlin S, Follenzi A, Shaffer SA, Stern LJ & Santambrogio L (2016) The dendritic cell major histocompatibility complex II (MHC II) Peptidome derives from a variety of processing pathways and includes peptides with a broad Spectrum of HLA-DM sensitivity. *J Biol Chem* 291, 5576–5595. [PubMed: 26740625]
41. Paziewska A, Polkowski M, Rubel T, Karczmarski J, Wiechowska-Kozłowska A, Dabrowska M, Mikula M, Dadlez M & Ostrowski J (2018) Mass spectrometry-based comprehensive analysis of pancreatic cyst fluids. *Biomed Res Int* 2018, 7169595. [PubMed: 30627566]

42. Deutsch EW, Omenn GS, Sun Z, Maes M, Pernemalm M, Palaniappan KK, Letunica N, Vandenbrouck Y, Brun V, Tao SC et al. (2021) Advances and utility of the human plasma proteome. *J Proteome Res* 20, 5241–5263. [PubMed: 34672606]
43. Sandy JD, Flannery CR, Neame PJ & Lohmander LS (1992) The structure of aggrecan fragments in human synovial fluid. Evidence for the involvement in osteoarthritis of a novel proteinase which cleaves the Glu 373-ala 374 bond of the interglobular domain. *J Clin Invest* 89, 1512–1516. [PubMed: 1569188]
44. Tortorella MD, Pratta M, Liu RQ, Austin J, Ross OH, Abbaszade I, Burn T & Arner E (2000) Sites of aggrecan cleavage by recombinant human aggrecanase-1 (ADAMTS-4). *J Biol Chem* 275, 18566–18573. [PubMed: 10751421]
45. Martin DR, Witten JC, Tan CD, Rodriguez ER, Blackstone EH, Pettersson GB, Seifert DE, Willard BB & Apte SS (2020) Proteomics identifies a convergent innate response to infective endocarditis and extensive proteolysis in vegetation components. *JCI Insight* 5,
46. Mustor EM, Frey AM, Jobson ME, Chaput D & Shaw LN (2024) Development of a new N-Terminomic method to study the Pathodegradome of the *Staphylococcus aureus* V8 protease in human neutrophils. *bioRxiv* 10, e0069725.
47. Frey AM, Chaput D & Shaw LN (2021) Insight into the human pathodegradome of the V8 protease from *Staphylococcus aureus*. *Cell Rep* 35, 108930. [PubMed: 33826899]
48. Gordon MH, Anowai A, Young D, Das N, Campden RI, Sekhon H, Myers Z, Mainoli B, Chopra S, Thuy-Boun PS et al. (2019) N-Terminomics/TAILS profiling of proteases and their substrates in ulcerative colitis. *ACS Chem Biol* 14, 2471–2483. [PubMed: 31393699]
49. Wichert R, Ermund A, Schmidt S, Schweinlin M, Ksiazek M, Arnold P, Knittler K, Wilkens F, Potempa B, Rabe B et al. (2017) Mucus detachment by host metalloprotease Meprin β requires shedding of its inactive pro-form, which is abrogated by the pathogenic protease RgpB. *Cell Rep* 21, 2090–2103. [PubMed: 29166602]
50. Cooper KN, Terekhova M, Potempa B, D'Aubeterre A, Timbol J, Chen SA, Dufour A, Jin S, Artyomov MN, Potempa J et al. (2025) The trapping of live neutrophils by macrophages during infection. *Cell Death Dis* 16, 488. [PubMed: 40610431]
51. Pablos I, Machado Y, de Jesus HCR, Mohamud Y, Kappelhoff R, Lindskog C, Vlok M, Bell PA, Butler GS, Grin PM et al. (2021) Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CL (pro) substrate degradome. *Cell Rep* 37, 109892. [PubMed: 34672947]
52. Pagotto A, Uliana F, Cavedon E, Nordio G, Pierangelini A, Acquasaliente L, Macchia ML, Bellanda M, Gatto B, De Silvestro G et al. (2025) The main protease (M(pro)) from SARS-CoV-2 triggers plasma clotting in vitro by activating coagulation factors VII and FXII. *Commun Biol* 8, 1145. [PubMed: 40750828]
53. Chaves AFA, de Barros B, Cosenza-Contreras M, Morone M, Sachetto ATA, Pinter N, Schmid M, Santoro ML, Schilling O & Serrano SMT (2025) Signatures of the systemic effects of a Snake venom and Antivenom: multiomics profiling of the kidney pathology. *Mol Cell Proteomics* 24, 101023. [PubMed: 40582410]
54. Bertholim L, Chaves AFA, Oliveira AK, Menezes MC, Asega AF, Tashima AK, Zelanis A & Serrano SMT (2021) Systemic effects of hemorrhagic snake venom metalloproteinases: untargeted peptidomics to explore the pathodegradome of plasma proteins. *Toxins (Basel)* 13, 764. doi: 10.3390/toxins13110764. [Correction added on 04 April 2026 after first online publication: This reference has been replaced, as the previous one was included in error.] [PubMed: 34822548]
55. Vázquez Torres S, Benard Valle M, Mackessy SP, Menzies SK, Casewell NR, Ahmadi S, Burlet NJ, Muratspahić E, Sappington I, Overath MD et al. (2025) De novo designed proteins neutralize lethal snake venom toxins. *Nature* 639, 225–231. [PubMed: 39814879]
56. Hartley B, Bassiouni W, Roczkowsky A, Fahlman R, Schulz R & Julien O (2024) Data-independent acquisition proteomics and N-Terminomics methods reveal alterations in mitochondrial function and metabolism in ischemic-Reperfused hearts. *J Proteome Res* 23, 844–856. [PubMed: 38264990]
57. Prudova A, Gocheva V, Auf dem Keller U, Eckhard U, Olson OC, Akkari L, Butler GS, Fortelny N, Lange PF, Mark JC et al. (2016) TAILS N-Terminomics and proteomics show protein degradation dominates over proteolytic processing by Cathepsins in pancreatic tumors. *Cell Rep* 16, 1762–1773. [PubMed: 27477282]

58. Ziegler AR, Parker BL, Scott NE & Edgington-Mitchell LE (2025) Employing complementary fractionation-based N-terminomics approaches enhances the identification of legumain cleavage events in naïve and inflamed colon tissue. *Protein Sci* 34, e70186. [PubMed: 40970443]
59. Song X, Ren X, Mei Q, Liu H & Huang H (2024) Advancing in-depth N-Terminomics detection with a cleavable 2-Pyridinecarboxyaldehyde probe. *J Am Chem Soc* 146, 6487–6492. [PubMed: 38421262]
60. Lorentzen LG, Yeung K, Zitkeviciute A, Yang-Jensen KC, Eldrup N, Eiberg JP & Davies MJ (2025) N-terminal proteomics reveals distinct protein degradation patterns in different types of human atherosclerotic plaques. *J Proteome Res* 24, 144–157. [PubMed: 39665830]
61. Sabino F, Egli FE, Savickas S, Holstein J, Kaspar D, Rollmann M, Kizhakkedathu JN, Pohlmann T, Smola H & Auf dem Keller U (2018) Comparative Degradomics of porcine and human wound exudates unravels biomarker candidates for assessment of wound healing progression in trauma patients. *J Invest Dermatol* 138, 413–422. [PubMed: 28899681]
62. Sabino F, Hermes O & Auf dem Keller U (2017) Body fluid Degradomics and characterization of basic N-Terminome. *Methods Enzymol* 585, 177–199. [PubMed: 28109429]
63. Wildes D & Wells JA (2010) Sampling the N-terminal proteome of human blood. *Proc Natl Acad Sci USA* 107, 4561–4566. [PubMed: 20173099]
64. Das N, de Almeida LGN, Derakhshani A, Young D, Mehdinejadani K, Salo P, Rezansoff A, Jay GD, Sommerhoff CP, Schmidt TA et al. (2023) Trypsin β regulation of joint lubrication and inflammation via proteoglycan-4 in osteoarthritis. *Nat Commun* 14, 1910. [PubMed: 37024468]
65. Bhutada S, Hoyle A, Piuze NS & Apte SS (2025) Degradomics defines proteolysis information flow from human knee osteoarthritis cartilage to matched synovial fluid and the contributions of secreted proteases ADAMTS5, MMP13 and CMA1 to articular cartilage breakdown. *Osteoarthritis Cartil* 33, 116–127.
66. Demir F, Kovalenko E, Lassé M, Svenningsen EB, Bernth Jensen JM, Billing AM, Groeneveld K, Hutzfeldt A, Nilges L, Guerra JPL et al. (2025) Proteolytic profiling of human plasma reveals an immunoactive complement C3 fragment. *EMBO J* 44, 7721–7758. [PubMed: 41145914]
67. Kapilan A, Bulluss M, Ziegler AR, Dabaja M, Derakhshani A, Anowai A, Armstrong V, Campden R, Young D, Sun YJ et al. (2025) N-terminomics and proteomics analysis of Calpain-2 reveal key proteolytic processing of metabolic and cell adhesion proteins. *Protein Sci* 34, e70144. [PubMed: 40277457]
68. Nandadasa S, Martin D, Deshpande G, Robert KL, Stack MS, Itoh Y & Apte SS (2023) Degradomic identification of membrane type 1-matrix metalloproteinase as an ADAMTS9 and ADAMTS20 substrate. *Mol Cell Proteomics* 22, 100566. [PubMed: 37169079]
69. Gal J, Dufour A, Young D, Yang ES & Geddes JW (2025) The identification of proteolytic substrates of Calpain-5 with N-Terminomics. *Int J Mol Sci* 26,
70. Anderson BM, de Almeida LGN, Sekhon H, Young D, Dufour A & Edgington-Mitchell LE (2020) N-Terminomics/TAILS profiling of macrophages after chemical inhibition of Legumain. *Biochemistry* 59, 329–340. [PubMed: 31774660]
71. Tom I, Pham VC, Katschke KJ Jr, Li W, Liang WC, Gutierrez J, Ah Young A, Figueroa I, Eshghi ST, Lee CV et al. (2020) Development of a therapeutic anti-HtrA1 antibody and the identification of DKK3 as a pharmacodynamic biomarker in geographic atrophy. *Proc Natl Acad Sci USA* 117, 9952–9963. [PubMed: 32345717]
72. Schlage P, Egli FE, Nanni P, Wang LW, Kizhakkedathu JN, Apte SS & auf dem Keller U (2014) Time-resolved analysis of the matrix metalloproteinase 10 substrate degradome. *Mol Cell Proteomics* 13, 580–593. [PubMed: 24281761]
73. Plasman K, Van Damme P, Kaiserman D, Impens F, Demeyer K, Helsens K, Goethals M, Bird PI, Vandekerckhove J & Gevaert K (2011) Probing the efficiency of proteolytic events by positional proteomics. *Mol Cell Proteomics* 10, M110.003301.
74. O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR & Folkman J (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* 88, 277–285. [PubMed: 9008168]

75. An YA, Xiong W, Chen S, Bu D, Rutkowski JM, Berger JP, Kusminski CM, Zhang N, An Z & Scherer PE (2023) Endotrophin neutralization through targeted antibody treatment protects from renal fibrosis in a podocyte ablation model. *Mol Metab* 69, 101680. [PubMed: 36696925]
76. McCulloch DR, Nelson CM, Dixon LJ, Silver DL, Wylie JD, Lindner V, Sasaki T, Cooley MA, Argraves WS & Apte SS (2009) ADAMTS metalloproteases generate active versican fragments that regulate interdigital web regression. *Dev Cell* 17, 687–698. [PubMed: 19922873]
77. Lark MW, Bayne EK, Flanagan J, Harper CF, Hoerner LA, Hutchinson NI, Singer II, Donatelli SA, Weidner JR, Williams HR et al. (1997) Aggrecan degradation in human cartilage. Evidence for both matrix metalloproteinase and aggrecanase activity in normal, osteoarthritic, and rheumatoid joints. *J Clin Invest* 100, 93–106. [PubMed: 9202061]
78. Sandy JD, Westling J, Kenagy RD, Iruela-Arispe ML, Verscharen C, Rodriguez-Mazaneque JC, Zimmermann DR, Lemire JM, Fischer JW, Wight TN et al. (2001) Versican V1 proteolysis in human aorta in vivo occurs at the Glu441-Ala442 bond, a site that is cleaved by recombinant ADAMTS-1 and ADAMTS-4. *J Biol Chem* 276, 13372–13378. [PubMed: 11278559]
79. Matthews RT, Gary SC, Zerillo C, Pratta M, Solomon K, Arner EC & Hockfield S (2000) Brain-enriched hyaluronan binding (BEHAB)/brevican cleavage in a glioma cell line is mediated by a disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) family member. *J Biol Chem* 275, 22695–22703. [PubMed: 10801887]
80. Bay-Jensen AC, Kjølgaard-Petersen CF, Petersen KK, Arendt-Nielsen L, Quaschnick HL, Mobasheri A, Karsdal MA & Leeming DJ (2018) Aggrecanase degradation of type III collagen is associated with clinical knee pain. *Clin Biochem* 58, 37–43. [PubMed: 29702061]
81. Billingham RC, Buxton EM, Edwards MG, McGraw MS & McIlwraith CW (2001) Use of an antineoepitope antibody for identification of type-II collagen degradation in equine articular cartilage. *Am J Vet Res* 62, 1031–1109. [PubMed: 11453476]
82. Ahrman E, Lorenzo P, Holmgren K, Grodzinsky AJ, Dahlberg LE, Saxne T, Heinegard D & Onnerfjord P (2014) Novel cartilage oligomeric matrix protein (COMP) neoepitopes identified in synovial fluids from patients with joint diseases using affinity chromatography and mass spectrometry. *J Biol Chem* 289, 20908–20916. [PubMed: 24917676]
83. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR & Vandekerckhove J (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 21, 566–569. [PubMed: 12665801]
84. Venne AS, Vögler FN, Meisinger C, Sickmann A & Zahedi RP (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. [PubMed: 23964590]
85. Weng SSH, Demir F, Ergin EK, Dirnberger S, Uzozie A, Tuscher D, Nierves L, Tsui J, Huesgen PF & Lange PF (2019) Sensitive determination of proteolytic Proteoforms in limited microscale proteome samples. *Mol Cell Proteomics* 18, 2335–2347. [PubMed: 31471496]
86. Chen L, Shan Y, Weng Y, Sui Z, Zhang X, Liang Z, Zhang L & Zhang Y (2016) Hydrophobic tagging-assisted N-termini enrichment for in-depth N-Terminome analysis. *Anal Chem* 88, 8390–8395. [PubMed: 27532682]
87. Mommen GP, van de Waterbeemd B, Meiring HD, Kersten G, Heck AJ & de Jong AP (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. [PubMed: 22729381]
88. Mustor EM, Frey AM, Jobson ME, Chaput D & Shaw LN (2025) Development of a new N-terminomic method to study the pathodegradome of the *Staphylococcus aureus* V8 protease in human neutrophils. *mSystems* 10, e0069725. [PubMed: 40558068]
89. Morikawa K, Nishida H, Imami K & Ishihama Y (2024) One-step N-Terminomics based on isolation of protein N-terminal peptides from LysargiNase digests by tip-based strong cation exchange chromatography. *Mol Cell Proteomics* 23, 100820. [PubMed: 39069075]
90. Xu G, Shin SB & Jaffrey SR (2009) Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc Natl Acad Sci USA* 106, 19310–19315. [PubMed: 19892738]

91. Griswold AR, Cifani P, Rao SD, Axelrod AJ, Miele MM, Hendrickson RC, Kentsis A & Bachovchin DA (2019) A chemical strategy for protease substrate profiling. *Cell Chem Biol* 26, 901–907. [PubMed: 31006619]
92. Bridge HN, Leiter W, Frazier CL & Weeks AM (2024) An N terminomics toolbox combining 2-pyridinecarboxaldehyde probes and click chemistry for profiling protease specificity. *Cell Chem Biol* 31, 534–549. [PubMed: 37816350]
93. Tanco S, Aviles FX, Gevaert K, Lorenzo J & Van Damme P (2017) Identification of carboxypeptidase substrates by C-terminal COFRADIC. *Methods Mol Biol* 1574, 115–133. [PubMed: 28315247]
94. Chen L, Shan Y, Yang C, Sui Z, Zhang X, Zhang L & Zhang Y (2020) Carboxypeptidase B-assisted charge-based fractional diagonal chromatography for deep screening of C-Terminome. *Anal Chem* 92, 8005–8009. [PubMed: 32441514]
95. Schilling O, Huesgen PF, Barré O & Overall CM (2011) Identification and relative quantification of native and proteolytically generated protein C-termini from complex proteomes: C-terminome analysis. *Methods Mol Biol* 781, 59–69. [PubMed: 21877277]
96. Zhang Y, He Q, Ye J, Li Y, Huang L, Li Q, Huang J, Lu J & Zhang X (2015) Systematic optimization of C-terminal amine-based isotope labeling of substrates approach for deep screening of C-Terminome. *Anal Chem* 87, 10354–10361. [PubMed: 26361894]
97. Du X, Fang C, Yang L, Bao H, Zhang L, Yan G & Lu H (2019) In-depth analysis of C Terminomes based on LysC digestion and site-selective Dimethylation. *Anal Chem* 91, 6498–6506. [PubMed: 31025853]
98. Nishida H & Ishihama Y (2022) One-step isolation of protein C-terminal peptides from V8 protease-digested proteins by metal oxide-based ligand-exchange chromatography. *Anal Chem* 94, 944–951. [PubMed: 34962382]
99. Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng LY, Messner S, Ehrenberger T, Zanotelli V, Butscheid Y, Escher C et al. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* 14, 1400–1410. [PubMed: 25724911]
100. Tyanova S, Temu T & Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 11, 2301–2319. [PubMed: 27809316]
101. Pino LK, Searle BC, Bollinger JG, Nunn B, MacLean B & MacCoss MJ (2020) The skyline ecosystem: informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* 39, 229–244. [PubMed: 28691345]
102. Gupta S & Röst H (2021) Automated workflow for peptide-level quantitation from DIA/SWATH-MS data. *Methods Mol Biol* 2228, 453–468. [PubMed: 33950509]
103. Rydén M, Turkiewicz A, Önnérkjöld P, Tjörnstrand J, Englund M & Ali N (2023) Identification and quantification of degradome components in human synovial fluid reveals an increased proteolytic activity in knee osteoarthritis patients vs controls. *Proteomics* 23, e2300040. [PubMed: 37226369]
104. Wang H & Julien O (2023) CaspSites: a database and web application for experimentally observed human caspase substrates using N-Terminomics. *J Proteome Res* 22, 454–461. [PubMed: 36696595]

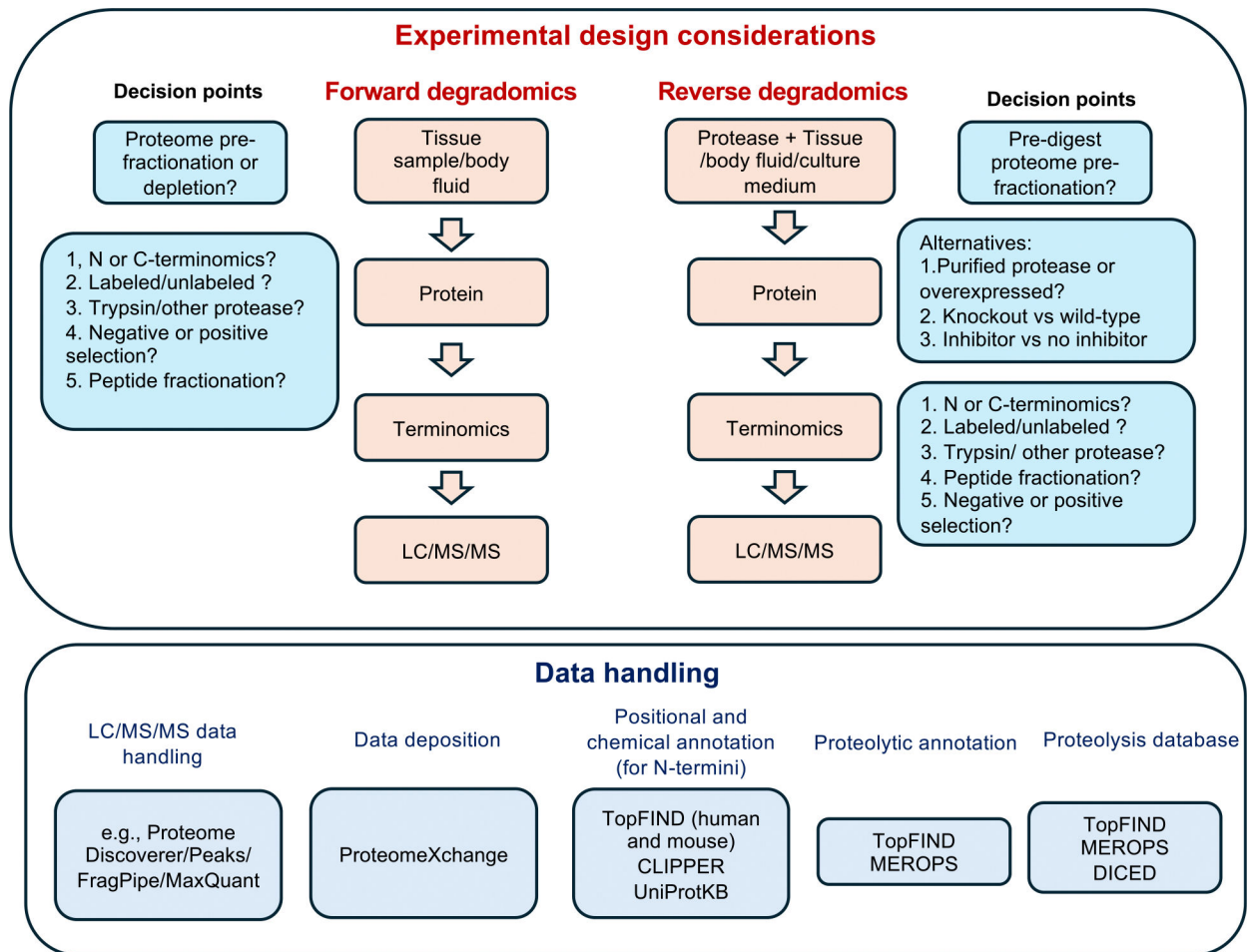


Fig. 1.
Overview of the experimental design considerations for forward and reverse degradomics with key decision points indicated and summary of the main data analysis aspects.

Table 1.

Degradomics workflows.

Workflow name	N-term/C-term	Enrichment	Label	Fractionation	Proteome digestion
TAILS [2]	N-term	Negative (HPG-ALD)	TMT, dimethyl		Trypsin
COFRADIC [83]	N-term	Negative		SCX	Trypsin
ChaFRADIC [84]	N-term	Negative		SCX	Trypsin
HUNTER [85]	N-term	Negative (Undecanal)	TMT, dimethyl		Trypsin
HYTANE [86]	N-term	Negative (hexadecanal)	Dimethyl, TMT		Trypsin
P-TAG [87]	N-term	Negative (TIO2 affinity)	TMT, dimethyl, GAP3		Trypsin
TAGS-CR [88]	N-term	Negative		SCX	Trypsin
CHAMP-N [89]	N-term	Negative		SCX	LysangiNase (TrypN)
FAIMS [58]	N-term	None	TMT, dimethyl	FAIMS	Trypsin
Biotinylation-subtiligase [63]	N-term	Positive (biotin-avidin)	Subtiligase		Trypsin
N-CLAP [90]	N-term	Positive(biotin-avidin)	PTC in TFA		Trypsin
CHOPS [91]	N-term	Positive (Streptavidin)	2PCA		Trypsin
Immuno-N-terminomics (miNterm)	N-term	Positive (TMT resin)	TMT		Trypsin
PICS2 [92]	N-term	Positive (neutravidin)	disulfide-2PCA		Trypsin
CHOPPER	N-term	Positive (neutravidin)	alkene-2PCA		Trypsin
C-COFRADIC [93]	C-term	Negative (amine-reactive butyrate)		SCX	Trypsin
CPB-ChaFRADIC [94]	C-term			SCX	Carboxypeptidase B
C-TAILS [95,96]	C-term	Negative (amine polynmer)	EDC		Trypsin
LysC digestion, dimethylation and enrichment [97]	C-term	Negative (NHS resin)	2PCA		LysC
PBC	C-term	Negative (Streptavidin)	2PCA		Trypsin
MOLEX chromatography [98]	C-term	Negative		MOLEX	Gluc (V8)

Table 2.

Resources for degradome and proteome data analysis.

Package	Function
Proteome discoverer	DDA and DIA data analysis for Thermo Fisher Mass spectrometers
Spectronaut [99]	DDA and DIA data analysis and can analyze data from various mass spectrometers
MaxQuant [100]	DDA and DIA data analysis and can analyze data from various mass spectrometers
Trans Proteomic Pipeline [29]	DDA and DIA data analysis and can analyze data from various mass spectrometers
FragPipe [21]	DDA and DIA data analysis and can analyze data from various mass spectrometers
DIA-NN [22]	DIA data analysis and can analyze data from various mass spectrometers
Skyline [101]	DDA and DIA data analysis and can analyze data from various mass spectrometers
OpenSwath [102]	DIA data analysis and can analyze data from various mass spectrometers
UniProtKB	Provides reference protein database validated protein sequences and structure, domains, PTMs, cellular compartment annotations, mapped alternate cleavage sites, natural processing sites, and isoform information.
MANTI [26]	Peptide annotation for terminomics using MaxQuant output.
Terminer [27]	Peptide annotation for terminomics using FragPipe output
CLIPPER [28,30]	Peptide annotation for terminomics using Proteome Discoverer and Spectronaut. It can also be manually customized for other softwares. Create visualizations and perform statistics.
Perseus [31]	Create visualizations and perform statistics. Requires manually curated dataset for terminomics.
Proteasy [103]	R package to retrieve data from MEROPS for substrates and proteases of interest.
MEROPS [33]	Protease and protease inhibitors database. Experimentally validated substrate and cleavage site information.
TopFind [34]	Integrated data from UniProt, MEROPS and community submissions from experimental terminomics datasets. Includes PathFINDER for pathway analysis predictions.
DICED [35]	Experimentally generated forward and reverse N-terminomics datasets for disease-specific protease activity patterns.
CapSites [104]	Database for experimentally generated human caspase cleavage sites.