Title

Multiplex substrate profiling by mass spectrometry for proteases

Journal

Methods in Enzymology, 682(Biological Chemistry 399 12 2018)

ISSN

0076-6879

Authors

Rohweder, Peter J
Jiang, Zhenze
Hurysz, Brianna M
et al.

Publication Date

2023

DOI

10.1016/bs.mie.2022.09.009

Peer reviewed

HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2023 May 22.

Published in final edited form as:

Methods Enzymol. 2023 ; 682: 375–411. doi:10.1016/bs.mie.2022.09.009.

Multiplex substrate profiling by mass spectrometry for proteases

Peter J. Rohweder[a], **Zhenze Jiang**[b], **Brianna M. Hurysz**[b], **Anthony J. O'Donoghue**[b,*], **Charles S. Craik**[a,*]

[a]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, United States

[b]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, United States

Abstract

Proteolysis is a central regulator of many biological pathways and the study of proteases has had a significant impact on our understanding of both native biology and disease. Proteases are key regulators of infectious disease and misregulated proteolysis in humans contributes to a variety of maladies, including cardiovascular disease, neurodegeneration, inflammatory diseases, and cancer. Central to understanding a protease's biological role, is characterizing its substrate specificity. This chapter will facilitate the characterization of individual proteases and complex, heterogeneous proteolytic mixtures and provide examples of the breadth of applications that leverage the characterization of misregulated proteolysis. Here we present the protocol of Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS), a functional assay that quantitatively characterizes proteolysis using a synthetic library of physiochemically diverse, model peptide substrates, and mass spectrometry. We present a detailed protocol as well as examples of the use of MSP-MS for the study of disease states, for the development of diagnostic and prognostic tests, for the generation of tool compounds, and for the development of protease-targeted drugs.

1. Introduction

Proteases are enzymes that irreversibly hydrolyze amide bonds in peptide and protein substrates. Proteases were originally thought to be protein-degrading enzymes, but they are now appreciated as post-translationally modifying enzymes. What distinguishes proteases from other classes of post-translationally modifying enzymes is their ability to catalyze a functionally irreversible reaction- amide bond hydrolysis. Because of the irreversible nature of this post-translational modification, protease activity is tightly regulated at various levels. Regulation of proteolytic activity is accomplished by changes in expression levels (Duffy et al., 2011; Isaacson, Martin Jensen, Subrahmanyam, & Ghandehari, 2017; Overall & López-Otín, 2002), zymogen conversion (Khan & James, 1998; Neurath & Walsh, 1976), the presence of interacting partners (Bode & Huber, 1992; Opoku-Nsiah et al., 2022), and localization/pH (Yadati, Houben, Bitorina, & Shiri-Sverdlov, 2020). The ultimate goal of

*Corresponding authors: ajodonoghue@health.ucsd.edu; charles.craik@ucsf.edu.

these multiple, orthogonal means of regulation is to tightly control proteolysis, enabling cells to respond to external and internal stimuli dynamically.

Proteases regulate a wide variety of biological processes in humans. The sequential cascade of zymogen conversion and subsequent proteolysis by a family of cysteine proteases called caspases is responsible for the initiation and execution of programmed cell death (Kumar, 2007; McIlwain, Berger, & Mak, 2013). The enzymes responsible for the control of blood coagulation consist of numerous serine proteases that participate in a complex cascade of zymogen activation and proteolysis (Furie & Furie, 1988). Different proteases are essential for adaptive immunity as the proteasome and various aminopeptidases are responsible for generation of MHC I peptides (Monaco, 1992; Rock, Farfán-Arribas, & Shen, 2010) while MHC II peptides are generated by the cathepsin family of proteases (Hsing & Rudensky, 2005). Proteases are also key regulators of tissue development and differentiation, with roles in processes such as bone morphogenesis (Ortega, Behonick, Stickens, & Werb, 2003) and mammary development (Green & Lund, 2005). In addition to these nuanced biological roles, proteases also contribute broad, non-specific degradative effects in the case of protein digestion (Whitcomb & Lowe, 2007) and protein quality control (Gottesman, Wickner, & Maurizi, 1997). Given the breadth of protease function in humans, dysregulation of proteolysis contributes to all hallmarks of cancer biology (Hanahan & Weinberg, 2000, 2011) and is a major contributor to a variety of other diseases (López-Otín & Bond, 2008). Proteases perform similar roles in other organisms and are main effectors of infectious disease and the ability to rapidly characterize emergent proteases is critical in the response to pandemics (Agbowuro, Huston, Gamble, & Tyndall, 2018; Sacco et al., 2020).

1.1. The need for protease profiling technologies

Central to understanding the diverse and nuanced roles of proteases in biological systems is the ability to biochemically characterize these enzymes. This characterization has largely focused on their substrate specificity toward endogenous and synthetic substrates. The detailed characterization of a protease's substrate specificity provides a better understanding of its mechanistic enzymology and biological role. This knowledge enables a variety of translational applications, including the development of tools to track protease activity and the intelligent design of inhibitors and protease-activated prodrugs. Advances in technology have greatly improved our ability to qualitatively and quantitatively uncover the substrate specificity of proteases. Every protease profiling technique offers its own unique advantages and disadvantages and the selection of the right technique for a research problem of interest is critical. In order to appreciate the motivation for the methodology presented herein, Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS), it is important to review other available methodologies and the unmet needs that inspired the design of MSP-MS.

1.2. Synthetic combinatorial libraries of fluorescent substrates

The development of fluorescent protease substrates allowed the quantification of substrate specificity in simple and rapid biochemical assays. In these fluorescent substrate-based methods, a peptide sequence of interest is attached to a fluorophore and the cleavage of the substrate potentiates the fluorescent signal, allowing turnover of substrate to be directly and quantitatively measured in simple plate reader assays. Despite the power of

these techniques, there are limitations from the requirement of the fluorescent reporter element which deviates from the native peptide structure that proteases evolved to recognize. One such technique is Positional Scanning Synthetic Combinatorial Libraries (PSSCL), where the fluorescent substrates contain a direct amide bond between the peptide and the C-terminal fluorophore (Harris et al., 2000; Ruggles, Fletterick, & Craik, 2004; Schneider & Craik, 2009). Cleavage of this bond significantly increases the fluorescence signal of the fluorophore. This method allows the quantification of substrate preference for each amino acid at each position of a peptide substrate in a simple biochemical assay. Despite the benefits of this technique, the presence of a peptide-fluorophore amide bond restricts the scope of the assay since only the N-terminal side of a scissile bond contains amino acids residues that can be diversified. This means the method does not work for carboxypeptidases which recognize their substrates via their C-terminus, or for endopeptidases with substrate recognition elements that occurs on the C-terminal side of the scissile bond. More recently, an analogous peptide-fluorophore system has been described which uses an N-terminal fluorophore, enabling the characterization of the C-terminal preference of a protease (Kuriki et al., 2018). These substrates allow the quantification of carboxypeptidase substrate specificity but do not work for aminopeptidases and endopeptidases that require substrate recognition on the N-terminal side of the scissile bond. A thematically similar but unique approach uses internally quenched fluorescent peptide substrates where a fluorophore and a quencher are attached to opposite ends of a peptide substrate (Poreba et al., 2017). Cleavage within the peptide backbone separates the quencher and the fluorophore physically and this loss of proximity potentiates the fluorescent signal. This approach removes the need for a direct fluorophore-peptide conjugation but still requires the presence of a fluorophore and a quencher at both termini and as such is not amenable to quantifying exopeptidase activity unless these fluorophore or quenchers are orientated in a substrate binding pocket that can accommodate bulky side chains. The inclusion of fluorophore-quencher pairs, which are typically hydrophobic, can also negatively affect the physiochemical properties of these peptide libraries. Additionally, since cleavage anywhere along the backbone results in fluorescent signal, it requires additional investigation to identify the site of cleavage when substrate specificity is not known. While the combination of these related techniques has expanded the scope of peptide-fluorophore libraries and enabled their use with all classes of proteases, none of these techniques work for all classes of proteases and the reliance on a fluorescent reporter restricts the scope as not all proteases tolerate the inclusion of these elements. The need to test the turnover of each member of a substrate library individually also significantly hinders the throughput of such methods.

1.3. Substrate phage-display based methods

The advent of phage display libraries enabled the generation of large libraries of potential peptide substrates ($>10^8$), expanding the substrate sequence space greatly compared to methods using endogenous proteins from cell lysates or synthetic combinatorial libraries (Matthews & Wells, 1993; Scholle et al., 2006). In these techniques, a library of potential peptide substrates is employed where these substrates are fused to a coat protein on phage and displayed on the surface of the phage where they are immobilized via a solid support. The library is exposed to a protease of interest and cleaved substrate phage is released, infected into *E.coli* and propagated in iterative rounds of enrichment. Highly cleaved

substrate phage pools are then enriched over time and their sequences can be identified through the library plasmid encoding their sequence. This expansion of substrate scope enabled the generation of more nuanced specificity information and the rapid identification of optimized substrates, but it also has limitations. Due to the requirement of substrate phage immobilization, peptide sequences are flanked by some immobilization handle, typically a peptide/protein tag or a biotinylation site, and the other side of the peptide substrate is fused to the phage coat protein, meaning both the N and C-termini of the peptide substrate are blocked. Because of these requirements, this method does not work for exopeptidases. Additionally, substrate sequences are typically identified by sequencing of the encoding library plasmid, meaning the specific site of cleavage is not directly identified, requiring additional follow-up experiments to confirm the site of cleavage when a target protease's substrate specificity is unknown or very broad. This is further complicated by the presence of flanking peptide sequences for immobilization or fusion to the coat protein, as cleavage at these positions can also lead to phage enrichment. Additionally, propagation of substrate phage in *E.coli* also results in biases, such as depletion of library members that are cleaved by endogenous, bacterial proteases or that are not readily expressed or stable in the bacterial host. Overall, this method is high throughput and samples a large diversity of substrates but is limited to endopeptidases and typically requires substantial follow-up experiments to confirm cleavage sites.

1.4. Mass spectrometry-based methods

The continued improvement of mass spectrometry has enabled the development of additional techniques which directly identify sites of cleavage through MS/MS based sequencing of peptide products (Agard & Wells, 2009; van den Berg & Tholey, 2012). Multidimensional fractionation and subsequent mass spectrometric analysis has enabled the direct identification of endogenous protease substrates and greatly expanded the throughput of mapping endogenous protease cleavage sites (Bredemeyer et al., 2004; Dix, Simon, & Cravatt, 2008; Uliana et al., 2021). Additional steps that modify the neo-N-termini generated by proteolytic cleavage have expanded these techniques, offering greater sensitivity and superior quantification (Ross et al., 2004; Wiita, Seaman, & Wells, 2014). Because these mass spectrometry-based techniques use native proteins without the need for affinity or fluorescent tags, they are able to overcome the limitations of substrate scope seen with fluorescent combinatorial libraries and phage display. These approaches can theoretically be used to characterize all classes of proteases, including endopeptidases, aminopeptidases, and carboxypeptidases, and they are able to rapidly identify endogenous substrates, streamlining the study of a protease's biological role. Despite these advantages, whole lysate-based mass spectrometric techniques suffer from complicated and time-intensive workflows. Additionally, the use of native proteins as substrates results in confounding effects from secondary structure and restricts theoretical substrate diversity to what is found in the proteome.

1.5. MSP-MS conception and design

The pros and cons of these existing methods inspired the design of the methodology presented in this chapter- Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS). MSP-MS is a simple and robust mass spectrometry-based method that can rapidly

and quantitatively characterize proteolytic activity using a library of synthetic, model peptides. Because MSP-MS uses a library of unmodified peptides, it can characterize both endopeptidase and exopeptidase (amino/carboxypeptidase) activities simultaneously, making it a universal method for the characterization of proteolysis (O'Donoghue et al., 2012). The method is carried out *in vitro,* making it amenable to a wide variety of sample types, including both purified proteases and complex proteolytic mixtures of interest from various origins. MSP-MS also has the benefit of remaining robust following systematic modulation of conditions such as pH, temperature, and the presence of inhibitors or activators of interest. The MSP-MS peptide library is exposed to a sample of interest containing one or more proteases, and cleavage sites are identified over time by LC-MS/MS. The use of LC-MS/MS as the detection method allows the direct quantification of these cleavages, enabling both the identification of substrates of interest and kinetic characterization in a single assay. Compared to whole lysate-based mass spectrometric techniques, the defined MSP-MS peptide library is compatible with simpler workflows and analysis pipelines while still maintaining high sequence diversity. Overall, MSP-MS is a rapid, simple, quantitative, and flexible method for the universal characterization of proteolysis in a wide variety of sample types.

1.6. Rational design of MSP-MS peptide library

MSP-MS was designed to produce comprehensive and quantitative proteolytic signatures. Central to this approach is the design of a substrate library which encompasses high sequence diversity with the smallest possible library size. A minimal library allows easier production, maintenance, and data analysis. The design of the library is based on the two-site hypothesis which posits recognition of two independent features on a substrate is typically sufficient for protease binding and subsequent cleavage (O'Donoghue et al., 2012). Following this logic, a library of decapeptide sequences was designed to encompass all neighbor (XY) and near neighbor (X*Y, X**Y) combinations of amino acids using a novel algorithm. Cysteine was omitted due to its propensity to form disulfide bonds and the oxidation prone methionine was replaced with norleucine (defined as lowercase n). A set of 76 peptides containing all amino acid pairs was constructed using this algorithm (Fig. 1, Table 1). To generate more diversity, each amino acid in this parent library was then swapped with a physiochemically distinct counterpart to afford two additional libraries each with 76 decapeptide sequences. To this core set of decapeptides, pairs of amino acids were added to the N and C-termini in order to incorporate additional substrate diversity for exopeptidases, bringing the peptide length to 14 amino acids in total (Table 2). The 14 amino acid size is sufficient for extended substrate binding while minimizing secondary structure. The end result of this design philosophy is a library of 228 14-mer peptides which encompass high sequence and physiochemical diversity, containing a total of 2,964 unique potential cleavage sites.

1.7. MSP-MS workflow

To perform MSP-MS, a sample containing a protease of interest is incubated with an equimolar mixture of all 228 peptides in the MSP-MS library. This protease-containing sample can be a pure protease or a complex mixture of proteases, such as cell or tissue extracts, conditioned media from microbial or mammalian cells, or patient-derived

biofluids. Assays can be performed at any pH or temperature, in any assay buffer that lacks detergent, and in the presence or absence of one or more protease inhibitors. The peptide library-protease mixture is incubated over time with aliquots of the reaction removed at multiple time points and quenched with denaturants such as urea or guanidium chloride. These samples are then desalted and prepared for LC-MS/MS analysis. The assay can be modified to increase through-put and quantitation by using isobaric tags and then samples are prepared for LC-MS/MS/MS analysis (Lapek et al., 2019). The raw data from the mass spectrometry experiments are searched against the parent library to identify full length and cleaved peptides. Cleaved peptides are matched to the parent peptide and the cleavage site is subsequently identified. Statistical and kinetic analysis of this dataset are performed to generate the extended substrate specificity of the protease-containing sample and to identify a list of highly cleaved substrates of interest (Fig. 2). The depth of this data allows for multiple types of analyses that are detailed below.

1.8. Validation of a predicted protease

When working with an uncharacterized or predicted protease, confirming peptidase activity is crucial. In cases where sample is precious and activity is unknown, selection of the assay conditions is critical. Since MSP-MS is amenable to proteases of all catalytic subclasses (serine, threonine, cysteine, metallo, aspartyl and glutamic) and all types of activities (endopeptidase, carboxypeptidase, aminopeptidase), it is a convenient assay for confirmation of enzyme activity when working with novel, uncharacterized proteases. Cleavage of peptides in the MSP-MS library over time can confirm whether a protein of interest has peptidase activity.

1.9. Identification and characterization of peptide substrates of interest

The simplest output from MSP-MS is a list of cleaved substrates. If the cleavage of a given substrate is specific enough such that the products accumulate over time without further degradation, the percent product formation as a function of time can be used to extrapolate catalytic efficiency, directly identifying and characterizing substrates in a single assay. The result of this most basic analysis is a list of highly cleaved substrates which can be used to rapidly generate highly sensitive probes to track the protease of interest’s activity in simple or complex samples.

1.10. Characterization of substrate specificity

From the list of library cleavages identified by LC-MS/MS, the global protease specificity can be identified by determining the prevalence of amino acids in the cleaved dataset relative to the parent library at each position along the substrate backbone. This global fingerprint is typically represented graphically as an iceLogo, as shown in Fig. 2 (Colaert et al., 2009). As proteases usually bind peptide substrates in close vicinity to the site of cleavage, extended substrate specificity is typically reported for the four amino acids at either side of the site of cleavage, also known as the scissile bond. As such, the iceLogo shown in Fig. 3 displays amino acid preference for an octameric substrate, where the peptide is shown from N to C terminus moving from left to right, and the scissile bond is in the center. Residues to the N-terminal side of the scissile bond are referred to as the “P side” and residues to the C-terminal side of the scissile bond are referred to as the “P′ side”

(P-*prime* side) (NH_2-P4-P3-P2-P1/P1′-P2'-P3′-P4′-COOH). The Y-axis of an iceLogo is a measurement of preference, typically percent difference or fold change. For iceLogo image presentation, all amino acids that are favored or disfavored at each position can be shown or only amino acids that are statistically significant can be shown. Favored residues are shown by increasing size in the positive Y axis-direction and disfavored residues are shown by increasing size in the negative Y axis-direction. Additionally, a meta-analysis of the sites of cleavage within the parent peptide library can be used to determine whether a protease has a preference for cleavage location along the peptide backbone. Cleavage exclusively at the N- or C-terminus indicates the presence of an exopeptidase (aminopeptidase or carboxypeptidase, respectively) and cleavages found more distal from the termini indicate endopeptidase activity. This global specificity profile can be done both for purified proteases and for complex mixtures.

1.11. Characterization of proteolysis in complex samples

When working with heterogeneous, complex samples such as conditioned media or patient samples, rapid identification of proteases of interest in those samples is crucial. Because MSP-MS occurs in a controllable, *in vitro* setting, properties of these complex mixtures can be modulated to expedite the elucidation of specific proteases of interest in the sample. By performing the MPS-MS experiment in the presence of subtype specific protease inhibitors such as E-64 (a cysteine protease inhibitor) and pepstatin A (an aspartyl protease inhibitor), the contribution of different protease subclasses can be determined. A decrease in the total number of observed cleavages paired with shifts in the substrate specificity following treatment with a protease inhibitor, can link a class of proteases to a specific activity observed in the biological sample. Additionally, modulation of buffer conditions like pH provides another line of evidence, linking specific proteolysis of interest to optimal buffer conditions. A combination of these modulations can take a global proteolytic signature from the complex mixture and rapidly identify characteristics of the causative proteases. When combined with additional follow-up experiments, such as western blots, proteomics, immunodepletion or nativePAGE separation, elucidation of proteases responsible for a specific biological role is possible.

2. General method and statistical analysis

2.1. Reagents

Protease: The protease sample can be a purified protease (native or recombinant) with a single active site, a protease complex with multiple active sites (e.g. proteasome) or a complex biological sample containing multiple proteases.

Substrate library: All peptides are stored separately as 5 mM stocks in −80 °C in either DMSO, ethanol, water or a mixture of these solvents. To make up the stock library, 5 μL of each peptide stock is mixed at equal volume with all other peptides to yield a final volume of 1.14 mL with each peptide at a concentration of 21.93 μM. 20 μL of this mixture is aliquoted into 57 low binding 0.6 mL tubes and stored at −80 °C until needed. A fresh library mixture is made up once per year. The sequences of each peptide are listed in Table 2.

Assay buffers and solvents: Use an optimized assay buffer for the target protease, if known. Detergents should be avoided whenever possible due downstream ionizing effects in the mass spectrometer that will decrease the signal of the peptides. If an optimized buffer is not known, some common assay buffers that may be used include:

- 20 mM Citrate Phosphate, pH 3.5, 100 mM NaCl
- 20 mM Citrate Phosphate, pH 5.5, 100 mM NaCl
- 20 mM Tris–HCl, pH 7.5, 100 mM NaCl
- 8 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl (phosphate buffered saline)
- 8.0 mM Na_2HPO_4, 1.47 mM KH_2PO_4, 138 mM NaCl, 2.7 mM KCl, 0.5 mM $MgCl_2$, 0.9 mM $CaCl_2$ (Dulbecco's phosphate buffered saline)

For sample quenching, desalting and LC-MS/MS, the following solvents are used

- 8 M Guanidine hydrochloride (GuHCl) (Sigma)
- 8 M Urea (Sigma)
- Acetonitrile (ACN), LC-MS grade (Thermo)
- Formic acid (FA), LC-MS grade (Thermo)
- Methanol, LC-MS grade (Thermo)
- Trifluoracetic acid (TFA), LC-MS grade (Thermo)
- Water, LC-MS grade (Thermo)
- Solvent A: 0.1% FA/H_2O or 0.1% TFA/H_2O
- Solvent B: 0.1% FA in 50% ACN/H_2O or 0.1% TFA in 50% ACN/H_2O
- Solvent C: 0.1% FA in 70% ACN/H_2O or 0.1% TFA in 70% ACN/H_2O

2.2. Equipment and materials

Standard lab equipment for biochemistry

- Analytical Empore C18 Disks (CDS)
- Assay and elution tubes. Recommend using MAXYMum Recovery microtubes (Axygen), 0.6 mL for assay and 1.5 mL for elution.
- 2.0 mL collection tubes for washes
- Bench top centrifuge
- Bench top sonicator
- Computer equipped with proteomics analysis software. We mainly uses PEAKS 8.5 (Bioinformatics Solutions Inc.) for data processing.

- Liquid chromatography-tandem mass spectrometer. We use a Q Exactive Mass Spectrometer (Thermo) equipped with an Ultimate 3000 HPLC (Thermo). All LC-MS/MS parameter settings described here are based on our equipment.
- pH test strips (Millipore)
- Vacuum evaporation system, commonly referred to as speedvac

2.3. Procedure

MSP-MS assay consists of four steps:

1. Protease digestion of peptide library
2. Solid phase extraction of peptides
3. LC-MS/MS analysis
4. Data analysis

MSP-MS assays are generally conducted in quadruplicate reactions.

Protease digestion of peptide library

1. Dilute protease to 2x of a desired concentration in assay buffer. The target protease can be tested with fluorescent substrates to find the optimal buffer and enzyme concentration for MSP-MS assay.
2. Sonicate peptide library for 15 min before diluting from 21.93 μM to 1 μM in assay buffer to make a 2x stock solution in 0.6 mL MAXYMum Recovery microtubes.
3. Add 40 μL of 8 M GuHCl or 8 M urea into multiple tubes, which will be used later for quenching the reactions at numerous time intervals.
4. To start the enzyme reaction, combine an equal volume of 2x protease and 2x peptide library such that the final concentration of each peptide is 0.5 μM.
5. At defined time intervals of incubation (e.g., 15, 60 and 240 min), remove 20 μL for each replicate assay and add it to a tube containing 40 μL of 8 M GuHCl or 8 M urea to denature the protein and thus quench the reaction. Control samples (0 min) are prepared by combining the 10 μL of 2X protease with 40 μL of quenching reagent and after 15 min, 10 μL of 2X peptide library is added. Control samples can also be prepared by inactivating the protease(s) with heat, an inhibitor, or by using an active site mutant.
6. Store samples in −80 °C until ready for desalting.

Solid phase extraction of peptides

1. Prepare C18 StageTips[23] from Empore C18 extraction disks. Alternatively, use commercial C18 desalting spin columns or tips.
2. Wet StageTips by adding 150 μL ACN to the C18 tips and centrifuge at 400 x *g* for 1 min in a 2 mL collection tube. At each step, ensure the C18 tip is not

submerged in the eluted liquid. This may require replacement with fresh elution tubes, or disposal of the eluted liquid.

3. Add 150 μL Solvent A to the C18 tips and centrifuge at 400 x *g* for 2 min. Centrifuge for additional time until all liquid has gone through the tips. Repeat for 2 more cycles.

4. Add 60 μL 2% TFA/H_2O or 2% FA/H_2O to each MSP-MS sample to acidify. Mix by inversion of tube. Use pH strip to confirm that the final pH < 3.0 as this will promote better binding of peptides to C18 matrix.

5. Load the entire contents of each quenched/acidified reaction tube to the C18 StageTips. Centrifuge at 400 x *g* for 4–6 min. If necessary, centrifuge for additional time until all liquid has gone through the tips.

6. Add 150 μL of Solvent A to the tips to wash. Centrifuge at 400 x *g* for 3–4 min. Repeat for 2 more cycles. Centrifuge until all liquid has gone through the tips.

7. Replace wash tubes with 1.5 mL MAXYMum recovery tubes. Add 150 μLSolvent B to C18 tips and centrifuge at 400 x *g* for 4 min (2 cycles). Add 150 μL Solvent C to the tips and centrifuge at 400 x *g* for 4 min (1 cycle). Centrifuge for additional time until all liquid has gone through the tips. Total elution volume for each sample is ~450 μL.

8. Place tubes in a SpeedVac to evaporate the solvent. Store dried samples at −80 °C until needed.

9. Just prior to LC-MS/MS analysis, hydrate each sample in 80 μL of Solvent A. Vortex well and transfer 40 μL to a 300 μL conical insert with an autosampler vial. Dry-down the remaining 40 μL in the sample tube and use as a back-up if replicate injections are needed.

10. Place the vial in the autosampler.

LC-MS/MS analysis—Inject 0.4 μg of total peptide (4 μL) from each MSP-MS sample into a Q-Exactive Mass Spectrometer (Thermo) equipped with an Ultimate 3000 HPLC. Peptides are separated by reverse phase chromatography on a C18 column (1.7 μm bead size, 75 μmx 25 cm, 65 °C) at a flow rate of 300 nL/min using a 60-min linear gradient from 5% to 30% B, with solvent A: 0.1% formic acid in water and solvent B: 0.1% formic acid in acetonitrile. Survey scans are recorded over a 150–2000 m/z range (70,000 resolutions at 200 m/z, AGC target 3×10^6, 100 ms maximum). MS/MS is performed in data-dependent acquisition mode with HCD fragmentation (28 normalized collision energy) on the 12 most intense precursor ions (17,500 resolutions at 200 m/z, AGC target 1×10^5, 50 ms maximum, dynamic exclusion 20 s).

Data analysis—Data are processed using PEAKS 8.5 (Bioinformatics Solutions Inc.). MS2 data are searched against the tetradecapeptide library sequences with decoy sequences in reverse order. A precursor tolerance of 20 ppm and 0.01 Da for MS2 fragments are defined. No protease digestion is specified. Data are filtered to 1% peptide level false discovery rates with the target-decoy strategy. Peptides are quantified with label free

quantification and data are normalized by Loess-G algorithm and filtered by 0.3 peptide quality (Chawade et al., 2014). Outliers from replicates are removed by Dixon's Q testing when there were at least 3 replicate values found out of the 4 replicates for each condition for every peptide. Missing and zero values are imputed with random normally distributed numbers in the range of the average of the smallest 5% of the data±SD. Cleaved peptide products are defined as those with intensity scores of eightfold or more above the peptide intensity scores in the inactive enzyme sample.

Code and Resource Availability—The requisite code for analyzing MSP-MS experiments is available publicly at https://github.com/briannahurysz/MSP-MS along with detailed instructions for its use. The requisite code and instructions may also be obtained by contacting the corresponding authors Charles Craik (Charles.Craik@UCSF.edu) or Anthony O'Donoghue (ajodonoghue@health.ucsd.edu).

Time Taken—Time for the whole procedure varies depending on the incubation step of the assay and data analysis time. If taking timepoints of less than 4 h, then one day is sufficient to complete the assay and desalt the samples. Desalted peptides are dried down on the speed vacuum overnight and ready for mass spectrometry analysis the following day.

Troubleshooting

1. If peptides under digested or over digested then adjust the enzyme concentration and/or timepoints selected
2. If detergent is important for the enzyme assay, then use mass spectrometry friendly detergents such as *N*-octyl-β-glucopyranoside or *N*-dodecyl β-D-maltoside.
3. If low peptide coverage is observed, this may be due to high protein or peptide levels in the enzyme sample that mask the signal. This is seen when assaying neat serum. The solution is to dilute the sample 50-fold or more and then leave the assay run for longer so that the dilute proteases have sufficient time to cleave peptides.
4. If low peptide count then pH may not acidic enough for good peptide binding to desalting column. Ensure that the molarity of the assay buffer is less than 25 mM as higher concentrations require more TFA or FA to reduce pH to <3.0.
5. If low peptide count then peptides are too hydrophilic and do not bind to desalting column. Solution is to try alternative desalting columns that are not C18-based.
6. Low intensity on chromatograms—increase the concentration of peptides injected by either increasing the volume of assay mix removed at each timepoint, or decreasing the volume of 0.1% TFA used to resuspend the sample before injection
7. Little difference between the first and last timepoint—The protease may be unstable or auto-degrade after the initial time point, thus the assay could be

repeated at shorter intervals using higher enzyme concentrations. Additionally, ensure the protease is inactivated by urea and/or GuHCl. If it is resistant then the reaction continues in the presence of the denaturants.

2.4 Example MSP-MS experiment

As an example of an MSP-MS experiment, we assayed the serine protease domain of recombinant human hepsin (R&D Systems, 4776-SE). This enzyme, is also known as TMPRSS1, and the full-length protein is a Type II membrane protein with an extracellular serine protease domain. The enzyme was activated as outlined in the product datasheet and assayed in 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$ at a final concentration of 1 μg/mL. The assay was performed in quadruplicate reaction tubes, and an aliquot of each reaction was removed after 15, 30 and 60 min incubation. Four replicate control reactions were set up that consisted of hepsin combined with 8 M urea prior to the addition of the peptide library. These control reactions were defined as the '0 min' timepoint. All 16 samples were desalted in parallel and subjected to LC-MS/MS analysis. Data analysis was performed as outlined above and the raw and filtered data can be found in Supplementary Information in the online version at https://doi.org/10.1016/bs.mie.2022.09.009.

After 15 min incubation, a total of 59 cleaved peptide bonds were quantified. These consisted of cleaved peptide products that had peptide intensity values increased by eightfold or more when compared to the intensity of the same peptide in the 0 min dataset (q value< 0.05). The number of cleaved peptide bonds increased to 94 after 30 min incubation and to 105 after 60 min incubation (Supplementary Information in the online version at https://doi.org/10.1016/bs.mie.2022.09.009).

We show an example of four peptides that are amongst the most efficiently cleaved by hepsin. VDYIEHKDQVRRMN was cleaved between R and M, where M corresponds to norleucine. The shorter C-terminal dipeptide product was not detected as the m/z was below the lower m/z cut-off limit of 150. However, the longer N-terminal 12-mer product was quantified at each of the timepoint and a progress curve was generated (Fig. 3A). KARSAFAEMWPDHG was cleaved between R and S and only the larger C-terminal fragment was quantified (Fig. 3B). Both of these peptide products increased linearly the course of the assay. In another example, we show cleavage of a substrate close to the center of the peptide, such that both the N-terminal and C-terminal fragments can be quantified. Here, both cleaved products ionize to similar levels and therefore the progress curves for both products overlay (Fig. 3C). In the final example, the peptide HWAFRSRYHGPLAH is cleaved at two sites, between R and S (Fig. 3D), and between R and Y (Fig. 3E). When comparing both images, it is clear that the concentration of the N-terminal product, HWAFR (Fig. 3D) and the C-terminal product YHGPHAL (Fig. 3E) increase over time while SRYHGPLAH (Fig. 3D) and HQAFRSR (Fig. 3E) accumulate more slowly. This is because the latter two peptides are each cleaved at the second site thereby releasing the SR dipeptide. The SR peptide cannot be detected due to the low m/z. In Fig. 3F, we show the degradation pattern of the starting substrate that yields 4 products initially and then 2 of these products are further degraded to HWAFR and YHGPLAH that can be detected and SR that is too short to be detected.

We next looked at the position within the 14-mer peptides were each cleavage occurred (Fig. 4A). We found that no peptides are cleaved between the 1st and 3rd amino acids while five peptides were cleaved between the 3rd and 4th amino acids with KAR*SAFAEMWPDHG shown in Fig. 3B being an example of one of these peptides. Cleavage occurred at each of the other positions within the 14-mer peptide library, and only one peptide was cleaved at the C-terminus between residues 13 and 14. These data show that hepsin is predominately an endopeptidase as the majority of peptides are cleaved toward the middle of the peptide, away from the N- and C- termini.

We next generated an iceLogo profile of the P4 to P4′ residues for each of the 59 peptide cleavage sites that were quantified after 15 min incubation (Fig. 4B). We predict that these are the most effectively cleaved substrates within our library. The profile shows that hepsin has a very strong preference for cleavage after Arg with 54 out of 59 peptides having Arg at P1. The remaining five cleaved peptides had Lys as the P1 residue. Beyond P1, norleucine (Nle, M in iceLogo) and Leu were significantly enriched in P3 while only Phe was significantly enriched in P2. On the primed side, Ser, Ile, Nle and Gly were significantly enriched in P1′, Gly in P2′ and Thr in P4′. All other amino acids shown above the midline in Fig. 4B were increased in frequency but with P value< 0.05. Amino acids below the midline, were either never found or found at a very low frequency. In summary, human hepsin is an endopeptidase with a strong preference for cleaving on the C-terminal side of Arg and has some additional preferences for other amino acids at the surrounding subsites.

3. Examples of MSP-MS in the literature

3.1. MSP-MS for the study and characterization of purified proteases

The most straightforward application of MSP-MS is for the characterization of a purified, single protease. MSP-MS analysis of a single protease will provide both the substrate specificity as well as a list of highly cleaved substrates which can be used for designing fluorogenic reporter substrates, inhibitors, imaging agents or linker sequences for protease-activated therapeutics.

One compelling example of this approach is the characterization of Granzyme B by Zhao and Bardine (Zhao et al., 2021) for the development of a Granzyme B-activated imaging agent. Granzyme B is a serine protease with endopeptidase activity that is a primary effector of immune-mediated cell killing where it is exocytosed by cytotoxic immune cells toward target cells in order to trigger apoptosis. In this example, Granzyme B was purified from human derived cytotoxic T-cell lymphocytes and characterized by MSP-MS. The MSP-MS experiment confirmed previous reports of Granzyme B specificity, indicating a strong preference for aspartic acid in the P1 position and proline in the P2 position. Additionally, MSP-MS allowed identification of an efficiently cleaved substrate, IEPD*VSQV which was used in the development of a restricted interaction peptide (RIP) probe. Upon cleavage of the full RIP probe by Granzyme B, the peptide undergoes a conformational change and is deposited on membranes. Accumulation of the probe in a tissue of interest is indicative of Granzyme B activity in that tissue and thus indicative of an active immune response. The authors used this probe to track Granzyme B activity *in vivo* in a variety of indications, including the treatment of multiple cancer models with checkpoint blockade immunotherapy

and in a model of pulmonary inflammation. The use of MSP-MS in this study enabled the rapid identification of a highly cleaved substrate, which was reformatted into a functional probe that serves as a readout of immune activation in a living animal. The specificity of this probe for Granzyme B is paramount to its success as a tool to study immune activity.

In a second example of the value of MSP-MS in the study of a purified protease is the characterization of Pd_dinase by Xu et al. (2018). Pd_dinase was a putative protease identified from the genome of *Parabacteroides distasonis*, a human gut commensal bacterium. Showing homology to C1B-like proteases, Pd_dinase was predicted to be a cysteine protease based on sequence. The putative enzyme was expressed and then characterized using MSP-MS. A strong preference for the cleavage of dipeptides from the N-terminus confirmed that Pd_dinase is a protease with aminopeptidase activity that favors N-terminal glycine residues. Using this specificity profile, a Gly-Arg-aminomethylcoumarin (Gly-Arg-AMC) fluorescent substrate was synthesized and confirmed to be rapidly cleaved by Pd_dinase ($k_{cat} = 1.25 \pm 0.002\ s^{-1}$, $K_M = 25.6 \pm 1.7\ \mu M$). This substrate was then converted into an inhibitor by the replacement of the AMC group with an acyloxymethyl ketone warhead to afford Gly-Arg-AOMK. This inhibitor was shown to irreversibly inactivate Pd_dinase ($K_i = 683 \pm 36$ nM, $k_{inact} = 3.6 \times 10^{-4} \pm 1.0 \times 10^{-5}\ s^{-1}$). The use of this inhibitor enabled the solving of both an apo and inhibitor-bound crystal structure, informing the mechanistic and structural basis of the observed activity. In this example, MSP-MS was used to confirm peptidase activity for a putative protease, rapidly characterize the substrate specificity, and aid in the design of both a fluorescent substrate and a covalent inhibitor for the protease Pd_dinase.

Another example of the use of MSP-MS in the study of a purified protease is the characterization of the pH dependence of Cathepsin B activity by Yoon et al. (2021). Cathepsin B is a lysosomally-associated cysteine protease with a predominant carboxypeptidase function and a relatively minor endopeptidase capability. Cathepsin B is often mislocalized to the cytoplasm or the extracellular space in disease, so the authors sought to understand whether the neutral pH in these compartments relative to the highly acidic lysosome might result in altered substrate specificity which could be exploited for pH-dependent inhibition of Cathepsin B. By performing MSP-MS at different pH conditions, the authors were able to generate global specificity profiles as a function of pH and found that Cathepsin B showed an increased preference for glutamic acid at P2 under acidic conditions (pH 4.6) and an increased preference for arginine at P2 under neutral conditions (pH 7.2). Cathepsin B favored basic residues, such as arginine and lysine, at the P1 position in both acidic and neutral pH conditions. Using this information, two substrates were synthesized and tested. One substrate, Z-Arg-Lys-AMC, showed a 6.5-fold increase in catalytic efficiency in neutral conditions (pH 7.2) relative to acidic conditions (pH 4.6), with 50% maximal activity occurring between pH 6.2 and pH 8.5. Another substrate, Z-Glu-Lys-AMC, showed 14.7-fold increase in catalytic efficiency in acidic conditions (pH 4.6) relative to neutral conditions (pH 7.2), with 50% maximal activity occurring between pH's 3.6 and 5.6. A third substrate, using a hydrophobic phenylalanine at the P2 position, Z-Phe-Arg-AMC, was cleaved with nearly equal efficiency (1.2x) in neutral conditions (pH 7.2) relative to acidic conditions (pH 4.6), with 50% maximal activity

occurring between pH's 3.8 and 8.6. The neutral pH-selective substrate was converted to an inhibitor through the addition of an acyloxymethyl ketone warhead. This novel inhibitor, Z-Arg-Lys-AOMK, was found to be 68-fold more selective for inhibition of Cathepsin B in neutral pH $\left(k_{inact}/K_i 1.1 \times 10^5\ M^{-1}\ s^{-1}\right)$ compared to acidic pH $\left(k_{inact}/K_i 1.8 \times 10^3\ M^{-1}\ s^{-1}\right)$. Z-Arg-Lys-AOMK is the first pH-selective inhibitor of Cathepsin B, paving the way for inhibitors which may be able to differentiate between pathogenic Cathepsin B that resides in the cytoplasm or extracellular space and non-pathogenic Cathepsin B the is a key component of lysosomes. In this example, MSP-MS conditions were modulated systematically with changes in pH, revealing highly nuanced and novel substrate specificities for Cathepsin B, enabling the design of pH-selective substrates and a pH-selective inhibitor.

Another example of the use of MPS-MS with purified proteases is the characterization of the human and the *Plasmodium falciparum* (the malaria-causing protozoa) proteasome by Li, et al (Li et al., 2016). The eukaryotic proteasome is a large, multi-component protease complex with three distinct proteolytic subunits responsible for the degradation of a large proportion of cellular proteins, playing a central role in maintaining homeostasis in many organisms. Due to its general and conserved role in homeostasis, it is an attractive target for drug development efforts focused on infectious disease, but selectively inhibiting a proteasome of pathogenic origin without disrupting the activity of the human proteasome has historically been challenging. Using MSP-MS, the authors characterized the substrate specificity of the human and malarial proteasome and found that the malarial proteasome showed an increased preference for tryptophan at both the P1 and P3 positions relative to the human proteasome. Systematic incorporation of tryptophan residues at the P1 and P3 positions along the canonical trileucine and vinyl sulfone (vs) scaffold found in common proteasome inhibitors such as MG132 and Z-L3-vs afforded the covalent inhibitors WLL-vs and WLW-vs. Using activity-based probes, the authors characterized their cross-reactivity with the three proteasome subunits (β1, β2, and β5) for both the malarial and human proteasome. WLL-vs was found to inhibit both the β2 and β5 subunit of the malarial proteasome but only the β5 subunit of the human proteasome. WLW-vs was found to inhibit only the β2 subunit of the malarial proteasome and only the β5 subunit of the human proteasome. Using the β2/β5 cross-reactive WLL-vs, the authors demonstrated that the inhibitor showed little toxicity to non-transformed human foreskin fibroblasts but was a potent inhibitor of *P. falciparum* cultures. Additionally, a single dose of WLL-vs resulted in an almost complete reduction of parasite burden in a *Plasmodium chabaudi* infected murine model as a single agent. The use of MSP-MS in this example enabled the identification of subtle differences in substrate specificity between evolutionarily similar proteases and enabled the design of two protease inhibitors with unique subunit and species-based activities which both show promise for the management of malarial infection.

3.2. MSP-MS for the study and characterization of complex proteolytic samples

The use of MSP-MS in complex, heterogenous samples offers the ability to profile global proteolysis in an unbiased manner. This approach can enable the identification of active proteases of interest in these samples and generate global "fingerprints" of proteolysis in these complex samples.

One example of the use of MSP-MS with complex proteolytic samples is the characterization of human neutrophil extracellular traps (NETs) by O'Donoghue et al. (2013). NETs are complex, heterogenous mixtures of antimicrobial factors exocytosed by neutrophils in webs of extracellular DNA. Using NETs isolated from three donor neutrophil samples, the authors characterized global proteolysis using MSP-MS and found that proteolysis was dominated by endopeptidase and carboxypeptidase activity with a strong preference for isoleucine, valine, and threonine at P1. Proteomics identified a set of candidate proteases responsible for this activity. In particular, neutrophil elastase (NE) was enriched in the stimulated NET samples compared to unstimulated controls. MSP-MS analysis of purified NE closely mirrored the experiments run with the heterogenous NETs, showing similar endopeptidase activity with a preference for isoleucine, valine, and threonine at P1. Additionally, of the 40 cleavages observed in NET samples from all three donor samples, 33 were shared with those of purified NE, indicating it was the dominant activity in these samples. NE was subsequently immune-depleted and the remining activity was attributed to Cathepsin G and Proteinase 3. Using MSP-MS, the authors were able to rapidly and comprehensively characterize proteolysis in a complex and poorly understood sample and confirm that proteolysis from NETs is dominated by neutrophil elastase activity, paving the way for future mechanistic studies on NET-based immune killing.

Another example of the use of MSP-MS to rapidly characterize heterogenous proteolytic samples is the work by Winter et al. to uncover biofilm-associated proteolysis in *Candida albicans* (Winter et al., 2016). By growing *Candida albicans* under both planktonic and biofilm conditions, the authors were able to generate conditioned media from both stages of the lifecycle and analyze their proteolytic signatures with MSP-MS. Overall, the samples from the biofilm were found to have more proteolysis (308 vs 185 cleaved sites), with 106 cleavages being shared between the two samples, 202 unique cleavages for the biofilm samples, and 79 unique cleavages for the planktonic samples. Proteomics of these samples identified Sap5 and Sap6 as proteases that might be responsible for the unique proteolysis observed in the biofilm samples. These proteases were expressed and purified. When subjected to MSP-MS analysis their specificities closely matched the biofilm proteolytic signature. Using the sequences of highly cleaved peptides found in the MSP-MS experiments with recombinant Sap5 and Sap6, fluorescent substrates were synthesized and used to measure proteolysis in planktonic and biofilm samples of WT *Candida albicans*. Selectivity for the substrates was confirmed using Sap5 and Sap6 knock out strains. In addition, turnover of the fluorescent substrates was much higher in WT biofilm samples relative to planktonic samples in eight different *Candida* strains, thereby confirming that these two proteases are broadly upregulated in *Candida* species during biofilm progression. Using MSP-MS, the authors were able to globally characterize proteolysis in *Candida albicans* samples in the planktonic and biofilm stages and identify proteolytic signatures that were specific for each. Using these signatures and those of purified, individual candidate proteases, they were able to identify two proteases important for biofilm progression across many *Candida* species and develop fluorescent substrates to track their activity.

Another application of MSP-MS for the study of a highly complex proteolytic sample is found in the work of Jiang, et al to characterize proteolysis in dense core secretory vesicles (DCSVs), a source of neuropeptides that modulate cell-cell signaling in the nervous and

endocrine system (Jiang et al., 2021). The authors characterized proteolysis in chromaffin granules (GC), a model system of DCSVs, with MSP-MS at pH 5.5 and 7.2 with various class-specific protease inhibitors. Proteolysis was found to be higher globally at pH 5.5 with 600 unique cleavages identified in the MSP-MS library compared to 204 at pH 7.2, with 148 of those cleavages being shared between the two conditions. The specificity profiles at both pH conditions showed similar but distinct fingerprints. Analysis of the cleavages suppressed with various protease inhibitors afforded specificity profiles for each class of protease present in the sample, revealing nuanced details including pepstatin-sensitive aspartic proteases were responsible for endopeptidase activity, E64-sensitive cysteine proteases were responsible for dicarboxypeptidase activity, and AEBSF-sensitive serine proteases were responsible for monocarboxypeptidase activity. With this information, purified single proteases identified from the proteomic profiling of GCs were further tested by MSP-MS. From this data, the authors were able to link certain activities identified in the GCs to individual proteases. The AEBSF-sensitive monocarboxypeptidase activity was found to correlate strongly to Cathepsin A. The E64-sensitive dicarboxypeptidase activity was found to correlate strongly to Cathepsin B. The pepstatin-sensitive endopeptidase activity was found to correlate strongly to Cathepsin D. Overall, 67.3% of the GC cleavages identified at pH 5.5 were attributable to the recombinant enzymes tested by the authors and 88.5% of the GC cleavages identified at pH 7.2 were attributable to these enzymes, confirming that Cathepsins A, B, C, D, and L are the dominant proteases in GCs. In this study, the authors were able to use MSP-MS to characterize complex proteolytic samples at multiple pH conditions and with multiple class-specific inhibitors in order to fully characterize proteolysis in these samples. Comparison of these inhibitor-sensitive recombinant protease profiles, coupled with proteomic data allowed for the direct identification of proteases of interest in these complex and dynamic proteolytic granules.

3.3. MSP-MS for the study and characterization of biofluids

Use of MSP-MS for the detection and characterization of proteolysis in biofluids has become a valuable application for the assay. It does not require any knowledge of the protease present in the sample and therefore facilitates unbiased discovery of active enzymes in a given sample. Characterization of biofluids taken directly from an organism or a patient allows a rapid means to characterize proteolysis in the highly complicated system. As no model system is as accurate as the parent organism itself, the ability to study proteolysis directly in these samples is of great value. Here we present examples of MSP-MS for the study of biofluids collected directly from whole organisms.

One interesting example of the use of MSP-MS for the study of biofluids is the work of Bibo-Verdugo, et al. to characterize proteolysis in the gastric juices of *Homarus americanus*, the American lobster (Bibo-Verdugo et al., 2016). Crustaceans are a diverse group adapted to varied environmental conditions, making them an interesting potential source of novel biology. Proteomic analysis of the gastric juices of *H. americanus* identified potential proteases Cathepsin L1, Cathepsin L2, Cathepsin L3, Cathepsin D1, and Cathepsin C2. MSP-MS analysis of gastric juices indicated the proteases present prefer phenylalanine, serine, and threonine at P1, and isoleucine, norleucine, and arginine at P1’. Treatment with the cysteine protease inhibitor E64 suppressed 92 of the 140 identified cleavages, indicating

that cysteine proteases were the dominant proteases in the sample. In order to more clearly elucidate the contribution of individual proteases, the gastric juice was fractionated on a native gel and individual bands were excised and incubated with the 228-member peptide library. These fractionated enzymes demonstrated unique activities. Of particular interest was the observation that Cathepsin L2 shows a strong preference for proline in the P2 position, in stark contrast to the closely related Cathepsin L3 which favors nearly all hydrophobic residues except proline in this position. Cathepsin D from *H. americanus* was further purified and compared to Cathepsin D from other sources, including porcine and bovine, and was found to have similar specificity. Kinetic characterization of Cathepsin D from these various sources was conducted with a commercial substrate closely resembling a highly cleaved, conserved substrate identified in the MSP-MS experiments. Cathepsin D from *H. americanus* was found to be substantially more tolerant of cold temperatures than Cathepsin D from bovine or porcine sources, corroborating the idea that enzymes from diverse environmental contexts have unique and favorable characteristics. Overall, the authors were able to characterize the proteolytic composition of *H. americanus* gastric juices without the guidance of an annotated genome using MSP-MS. MSP-MS revealed that closely related Cathepsin L proteins showed surprisingly divergent activities, enabling broad substrate scope with only a few, closely related enzymes.

One compelling example of the use of MSP-MS to study proteolysis in patient-derived samples is found in the work of Ivry, et al. to study misregulated proteolysis in precancerous pancreatic cysts (Ivry et al., 2017). Pancreatic cysts are present in 2–3% of the adult population, but a large proportion of these are benign. Currently, the ability to distinguish between benign, non-mucinous cysts and potentially precancerous, mucinous cysts is lacking and consequently the preoperative diagnosis of mucinous cysts is incorrect in up to 30% of cases. This results in the unnecessary resection of benign cysts and increased morbidity. Using samples collected directly from patients undergoing surgical resection with pathologically confirmed diagnoses, the authors characterized misregulated proteolysis in these samples using MSP-MS. As misregulated proteolysis is a hallmark of cancer, the authors hypothesized that a mucinous-specific proteolytic profile could enable the development of superior diagnostic tests. Global proteolysis was characterized both at pH 7.5 and pH 3.5 for a group of mucinous and non-mucinous cysts. Overall, it was found that proteolysis was largely conserved between the two groups at neutral pH, but a characteristic mucinous activity was revealed at acidic pH. Quantification of cleavages in the library using spectral counts allowed for fold-enrichment of individual cleavages across the two patient populations to be characterized and from this data, a specificity profile of cleavages enriched in mucinous samples was constructed. The MSP-MS assay was repeated at acidic pH in the presence of various, class-specific inhibitors and the mucinous-specific cleavages were found to be highly sensitive to pepstatin A, indicating that aspartyl proteases were largely responsible for this unique activity. Shotgun proteomics of mucinous and non-mucinous samples identified the presence of Cathepsin D, Cathepsin E, and Gastricsin. Quantification using spectral counts indicated that Cathepsin D was similarly present in mucinous and non-mucinous samples, but Cathepsin E and Gastricsin were enriched in mucinous samples. Immunohistochemistry analysis of Cathepsin E and Gastricsin confirmed their presence in the mucinous patient samples. Using MSP-MS with recombinant Gastricsin, the authors

identified a highly cleaved substrate that was selective for Gastricsin over Cathepsin D and Cathepsin E and synthesized it as an internally quenched fluorescent substrate. Using this Gastricsin-selective substrate and a previously reported Cathepsin E-selective substrate, the authors showed that the activity of these two proteases was significantly enriched in mucinous samples (n = 71) over non-mucinous samples (n = 39). With these two substrates as the basis of an activity-based diagnostic, the authors demonstrated that Cathepsin E activity was 92% specific and 70% sensitive for the diagnosis of mucinous cysts, and Gastricsin activity was 100% specific and 93% sensitive, outperforming the current clinical standard diagnostic CEA, which has a 94% specificity and 65% sensitivity for the diagnosis of mucinous cysts. A combination of Gastricsin activity and CEA resulted in a diagnostic with 100% specificity and 98% sensitivity for the diagnosis of mucinous cysts. Therefore, using MSP-MS, the authors were able to rapidly profile global proteolysis in patient-derived samples at multiple pH conditions and with various class-specific inhibitors in order to identify proteolytic activity that was characteristic of precancerous lesions. The resulting diagnostic assay is highly specific and sensitive and works with small amounts of patient-derived samples that can be readily collected via endoscopic ultrasound with fine needle aspiration.

4. Conclusions and perspectives

In this chapter we have described the technique Multiplex Substrate Profiling by Mass Spectrometry, or MSP-MS, for the characterization of proteolysis in a wide variety of samples and for a wide variety of applications. MSP-MS was designed to be a universal assay for the characterization of any type of proteolysis in any type of sample. We have described the use of MSP-MS for the characterization of purified proteases, including endopeptidases, aminopeptidases, and carboxypeptidases of various subtypes; we have described the use of MSP-MS for the characterization of complex proteolytic samples, including those derived both from human and non-human sources, and we have described the use of MSP-MS for the characterization of specimens collected directly from whole organisms, including patient samples collected in a clinical setting. Ultimately, MSP-MS is a technique for the study of proteolysis and the discovery of individual proteases of interest. Within these examples are various applications of the information garnered from MSP-MS, including the design of specific substrates for biochemical, prognostic, and diagnostic applications, and the design of specific inhibitors for therapeutic applications. As our understanding of the biological role of proteases continues to be expanded and becomes increasingly nuanced, techniques and methodologies that allow for the rapid, quantitative, flexible, and functional characterization of proteolysis are increasingly valuable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

A.J. O’Donoghue acknowledges the NIAID (R01AI158612) and the NCI (R21CA256460). C.S. Craik acknowledges the NIAID (U19AI171110). Brianna Hurysz was supported in part by the University of California-

San Diego (UCSD) Graduate Training Program in Cellular and Molecular Pharmacology through an institutional training grant from the National Institute of General Medical Sciences, T32GM007752.

References

Agard NJ, & Wells JA (2009). Methods for the proteomic identification of protease substrates. Current Opinion in Chemical Biology, 13(5–6), 503–509. 10.1016/j.cbpa.2009.07.026. [PubMed: 19729334]

Agbowuro AA, Huston WM, Gamble AB, & Tyndall JDA (2018). Proteases and protease inhibitors in infectious diseases. Medicinal Research Reviews, 38(4), 1295–1331. 10.1002/med.21475. [PubMed: 29149530]

Bibo-Verdugo B, O'Donoghue AJ, Rojo-Arreola L, Craik CS, & García-Carreño F (2016). Complementary proteomic and biochemical analysis of peptidases in lobster gastric juice uncovers the functional role of individual enzymes in food digestion. Marine Biotechnology, 18(2), 201–214. 10.1007/s10126-015-9681-5. [PubMed: 26613762]

Bode W, & Huber R (1992). Natural protein proteinase inhibitors and their interaction with proteinases. European Journal of Biochemistry, 204(2), 433–451. 10.1111/j.1432-1033.1992.tb16654.x. [PubMed: 1541261]

Bredemeyer AJ, Lewis RM, Malone JP, Davis AE, Gross J, Townsend RR, et al. (2004). A proteomic approach for the discovery of protease substrates. Proceedings of the National Academy of Sciences of the United States of America, 101(32), 11785–11790. 10.1073/pnas.0402353101. [PubMed: 15280543]

Chawade A, Alexandersson E, & Levander F (2014). Normalyzer: A tool for rapid evaluation of normalization methods for omics data sets. Journal of Proteome Research, 13(6), 3114–3120. 10.1021/pr401264n. [PubMed: 24766612]

Colaert N, Helsens K, Martens L, Vandekerckhove J, & Gevaert K (2009). Improved visualization of protein consensus sequences by iceLogo. Nature Methods, 6(11), 786–787. 10.1038/nmeth1109-786. [PubMed: 19876014]

Dix MM, Simon GM, & Cravatt BF (2008). Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell, 134(4), 679–691. 10.1016/j.cell.2008.06.038. [PubMed: 18724940]

Duffy MJ, Mullooly M, O'Donovan N, Sukor S, Crown J, Pierce A, et al. (2011). The ADAMs family of proteases: New biomarkers and therapeutic targets for cancer? Clinical Proteomics, 8(1), 9. 10.1186/1559-0275-8-9. [PubMed: 21906355]

Furie B, & Furie BC (1988). The molecular basis of blood coagulation. Cell, 53(4), 505–518. 10.1016/0092-8674(88)90567-3. [PubMed: 3286010]

Gottesman S, Wickner S, & Maurizi MR (1997). Protein quality control: Triage by chaperones and proteases. Genes & Development, 11(7), 815–823. 10.1101/gad.11.7.815. [PubMed: 9106654]

Green KA, & Lund LR (2005). ECM degrading proteases and tissue remodelling in the mammary gland. BioEssays, 27(9), 894–903. 10.1002/bies.20281. [PubMed: 16108064]

Hanahan D, & Weinberg RA (2000). The hallmarks of cancer. Cell, 100(1), 57–70. 10.1016/S0092-8674(00)81683-9. [PubMed: 10647931]

Hanahan D, & Weinberg RA (2011). Hallmarks of cancer: The next generation. Cell, 144(5), 646–674. 10.1016/j.cell.2011.02.013. [PubMed: 21376230]

Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, & Craik CS (2000). Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. Proceedings of the National Academy of Sciences of the United States of America, 97(14), 7754–7759. 10.1073/pnas.140132697. [PubMed: 10869434]

Hsing LC, & Rudensky AY (2005). The lysosomal cysteine proteases in MHC class II antigen presentation. Immunological Reviews, 207(1), 229–241. 10.1111/j.0105-2896.2005.00310.x. [PubMed: 16181340]

Isaacson KJ, Martin Jensen M, Subrahmanyam NB, & Ghandehari H (2017). Matrix-metalloproteinases as targets for controlled delivery in cancer: An analysis of upregulation and

expression. Journal of Controlled Release, 259, 62–75. 10.1016/j.jconrel.2017.01.034. [PubMed: 28153760]

Ivry SL, Sharib JM, Dominguez DA, Roy N, Hatcher SE, Yip-Schneider MT, et al. (2017). Global protease activity profiling provides differential diagnosis of pancreatic cysts. Clinical Cancer Research, 23(16), 4865–4874. 10.1158/1078-0432.CCR-16-2987. [PubMed: 28424202]

Jiang Z, Lietz CB, Podvin S, Yoon MC, Toneff T, Hook V, et al. (2021). Differential neuropeptidomes of dense core secretory vesicles (DCSV) produced at intravesicular and extracellular pH conditions by proteolytic processing. ACS Chemical Neuroscience, 12(13), 2385–2398. 10.1021/acschemneuro.1c00133. [PubMed: 34153188]

Khan AR, & James MNG (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Science, 7(4), 815–836. 10.1002/pro.5560070401. [PubMed: 9568890]

Kumar S (2007). Caspase function in programmed cell death. Cell Death & Differentiation, 14(1), 32–43. 10.1038/sj.cdd.4402060. [PubMed: 17082813]

Kuriki Y, Kamiya M, Kubo H, Komatsu T, Ueno T, Tachibana R, et al. (2018). Establishment of molecular design strategy to obtain activatable fluorescent probes for carboxypeptidases. Journal of the American Chemical Society, 140(5), 1767–1773. 10.1021/jacs.7b11014. [PubMed: 29368925]

Lapek JD, Jiang Z, Wozniak JM, Arutyunova E, Wang SC, Lemieux MJ, et al. (2019). Quantitative multiplex substrate profiling of peptidases by mass spectrometry. Molecular & Cellular Proteomics: MCP, 18(5), 968–981. 10.1074/mcp.TIR118.001099. [PubMed: 30705125]

Li H, O'Donoghue AJ, van der Linden WA, Xie SC, Yoo E, Foe IT, et al. (2016). Structure- and function-based design of Plasmodium-selective proteasome inhibitors. Nature, 530(7589), 233–236. 10.1038/nature16936. [PubMed: 26863983]

López-Otín C, & Bond JS (2008). Proteases: Multifunctional enzymes in life and disease. The Journal of Biological Chemistry, 283(45), 30433–30437. 10.1074/jbc.R800035200. [PubMed: 18650443]

Matthews DJ, & Wells JA (1993). Substrate phage: Selection of protease substrates by monovalent phage display. Science, 260(5111), 1113–1117. 10.1126/science.8493554. [PubMed: 8493554]

McIlwain DR, Berger T, & Mak TW (2013). Caspase functions in cell death and disease. Cold Spring Harbor Perspectives in Biology, 5(4), a008656. 10.1101/cshperspect.a008656. [PubMed: 23545416]

Monaco JJ (1992). A molecular model of MHC class-I-restricted antigen processing. Immunology Today, 13(5), 173–179. 10.1016/0167-5699(92)90122-N. [PubMed: 1386516]

Neurath H, & Walsh KA (1976). Role of proteolytic enzymes in biological regulation (a review). Proceedings of the National Academy of Sciences of the United States of America, 73(11), 3825–3832. [PubMed: 1069267]

O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, et al. (2012). Global identification of peptidase specificity by multiplex substrate profiling. Nature Methods, 9(11), 1095–1100. 10.1038/nmeth.2182. [PubMed: 23023596]

O'Donoghue AJ, Jin Y, Knudsen GM, Perera NC, Jenne DE, Murphy JE, et al. (2013). Global substrate profiling of proteases in human neutrophil extracellular traps reveals consensus motif predominantly contributed by elastase. PLoS One, 8(9), e75141. 10.1371/journal.pone.0075141. [PubMed: 24073241]

Opoku-Nsiah KA, de la Pena AH, Williams SK, Chopra N, Sali A, Lander GC, et al. (2022). The YΦ motif defines the structure-activity relationships of human 20S proteasome activators. Nature Communications, 13(1), 1226. 10.1038/s41467-022-28864-x.

Ortega N, Behonick D, Stickens D, & Werb Z (2003). How proteases regulate bone morphogenesis. Annals of the New York Academy of Sciences, 995(1), 109–116. 10.1111/j.1749-6632.2003.tb03214.x. [PubMed: 12814943]

Overall CM, & López-Otín C (2002). Strategies for MMP inhibition in cancer: Innovations for the post-trial era. Nature Reviews Cancer, 2(9), 657–672. 10.1038/nrc884. [PubMed: 12209155]

Poreba M, Szalek A, Rut W, Kasperkiewicz P, Rutkowska-Wlodarczyk I, Snipas SJ, et al. (2017). Highly sensitive and adaptable fluorescence-quenched pair discloses the substrate specificity

profiles in diverse protease families. Scientific Reports, 7(1), 43135. 10.1038/srep43135. [PubMed: 28230157]

Rock KL, Farfán-Arribas DJ, & Shen L (2010). Proteases in MHC Class I presentation and cross-presentation. The Journal of Immunology, 184(1), 9–15. 10.4049/jimmunol.0903399. [PubMed: 20028659]

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & Cellular Proteomics, 3(12), 1154–1169. 10.1074/mcp.M400129-MCP200. [PubMed: 15385600]

Ruggles SW, Fletterick RJ, & Craik CS (2004). Characterization of structural determinants of granzyme B reveals potent mediators of extended substrate specificity. The Journal of Biological Chemistry, 279(29), 30751–30759. 10.1074/jbc.M400949200. [PubMed: 15123647]

Sacco MD, Ma C, Lagarias P, Gao A, Townsend JA, Meng X, et al. (2020). Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. Science. Advances, 6(50), eabe0751. 10.1126/sciadv.abe0751. [PubMed: 33158912]

Schneider EL, & Craik CS (2009). Positional scanning synthetic combinatorial libraries for substrate profiling. Methods in Molecular Biology (Clifton, N.J.), 539, 59–78. 10.1007/978-1-60327-003-8_4.

Scholle MD, Kriplani U, Pabon A, Sishtla K, Glucksman MJ, & Kay BK (2006). Mapping protease substrates by using a biotinylated phage substrate library. ChemBioChem, 7(5), 834–838. 10.1002/cbic.200500427. [PubMed: 16628754]

Uliana F, Vizovišek M, Acquasaliente L, Ciuffa R, Fossati A, Frommelt F, et al. (2021). Mapping specificity, cleavage entropy, allosteric changes and substrates of blood proteases in a high-throughput screen. Nature Communications, 12(1), 1693. 10.1038/s41467-021-21754-8.

van den Berg BHJ, & Tholey A (2012). Mass spectrometry-based proteomics strategies for protease cleavage site identification. Proteomics, 12(4–5), 516–529. 10.1002/pmic.201100379. [PubMed: 22246699]

Whitcomb DC, & Lowe ME (2007). Human pancreatic digestive enzymes. Digestive Diseases and Sciences, 52(1), 1–17. 10.1007/s10620-006-9589-z. [PubMed: 17205399]

Wiita AP, Seaman JE, & Wells JA (2014). Global analysis of cellular proteolysis by selective enzymatic labeling of protein N-termini. Methods in Enzymology, 544, 327–358. 10.1016/B978-0-12-417158-9.00013-3. [PubMed: 24974296]

Winter MB, Salcedo EC, Lohse MB, Hartooni N, Gulati M, Sanchez H, et al. (2016). Global identification of biofilm-specific proteolysis in Candida albicans. MBio, 7(5), e01514–e01516. 10.1128/mBio.01514-16. [PubMed: 27624133]

Xu JH, Jiang Z, Solania A, Chatterjee S, Suzuki B, Lietz CB, et al. (2018). A commensal dipeptidyl aminopeptidase with specificity for N-terminal glycine degrades human-produced antimicrobial peptides in vitro. ACS Chemical Biology, 13(9), 2513–2521. 10.1021/acschembio.8b00420. [PubMed: 30085657]

Yadati T, Houben T, Bitorina A, & Shiri-Sverdlov R (2020). The ins and outs of cathepsins: Physiological function and role in disease management. Cells, 9(7), 1679. 10.3390/cells9071679. [PubMed: 32668602]

Yoon MC, Solania A, Jiang Z, Christy MP, Podvin S, Mosier C, et al. (2021). Selective neutral pH inhibitor of cathepsin B designed based on cleavage preferences at cytosolic and lysosomal pH conditions. ACS Chemical Biology, 16(9), 1628–1643. 10.1021/acschembio.1c00138. [PubMed: 34416110]

Zhao N, Bardine C, Lourenço AL, Wang Y, Huang Y, Cleary SJ, et al. (2021). In vivo measurement of granzyme proteolysis from activated immune cells with PET. ACS Central Science, 7(10), 1638–1649. 10.1021/acscentsci.1c00529. [PubMed: 34729407]

Bibliography of MSP-MS literature

Beekman C, Jiang Z, Suzuki BM, Palmer JM, Lindner DL, O'Donoghue AJ, et al. (2018). Characterization of PdCP1, a serine carboxypeptidase from Pseudogymnoascus destructans,

the causal agent of White-nose Syndrome. Biological Chemistry, 399(12), 1375–1388. 10.1515/hsz-2018-0240. [PubMed: 30367778]

Clarke SC, Dumesic PA, Homer CM, O'Donoghue AJ, Greca FL, Pallova L, et al. (2016). Integrated activity and genetic profiling of secreted peptidases in Cryptococcus neoformans reveals an aspartyl peptidase required for low pH survival and virulence. PLoS Pathogens, 12(12), e1006051. 10.1371/journal.ppat.1006051. [PubMed: 27977806]

Corvo I, Ferraro F, Merlino A, Zuberbühler K, O'Donoghue AJ, Pastro L, et al. (2018). Substrate specificity of cysteine proteases beyond the S2 pocket: Mutagenesis and molecular dynamics investigation of Fasciola hepatica Cathepsins L. Frontiers in Molecular Biosciences, 5. 10.3389/fmolb.2018.00040.

Corvo I, O'Donoghue AJ, Pastro L, Pi-Denis N, Eroy-Reveles A, Roche L, et al. (2013). Dissecting the active site of the collagenolytic cathepsin L3 protease of the invasive stage of Fasciola hepatica. PLoS Neglected Tropical Diseases, 7(7), e2269. 10.1371/journal.pntd.0002269. [PubMed: 23875031]

Dvořák J, Fajtová P, Ulrychová L, Leontovyč A, Rojo-Arreola L, Suzuki BM, et al. (2016). Excretion/secretion products from Schistosoma mansoni adults, eggs and schistosomula have unique peptidase specificity profiles. Biochimie, 122, 99–109. 10.1016/j.biochi.2015.09.025. [PubMed: 26409899]

Goupil LS, Ivry SL, Hsieh I, Suzuki BM, Craik CS, O'Donoghue AJ, et al. (2016). Cysteine and aspartyl proteases contribute to protein digestion in the gut of freshwater planaria. PLoS Neglected Tropical Diseases, 10(8), e0004893. 10.1371/journal.pntd.0004893. [PubMed: 27501047]

Guo Y, Wang Y, O'Donoghue AJ, Jiang Z, Carballar-Lejarazú R, Liang G, et al. (2020). Engineering of multiple trypsin/chymotrypsin sites in Cry3A to enhance its activity against Monochamus alternatus hope larvae. Pest Management Science, 76(9), 3117–3126. 10.1002/ps.5866. [PubMed: 32323409]

Ivry SL, Knudsen GM, Caiazza F, Sharib JM, Jaradeh K, Ravalin M, et al. (2019). The lysosomal aminopeptidase tripeptidyl peptidase 1 displays increased activity in malignant pancreatic cysts. Biological Chemistry, 400(12), 1629–1638. 10.1515/hsz-2019-0103. [PubMed: 31256057]

Ivry SL, Meyer NO, Winter MB, Bohn MF, Knudsen GM, O'Donoghue AJ, et al. (2018). Global substrate specificity profiling of post-translational modifying enzymes. Protein Science: A Publication of the Protein Society, 27(3), 584–594. 10.1002/pro.3352. [PubMed: 29168252]

Jones SK, Clarke SC, Craik CS, & Bennett RJ (2015). Evolutionary selection on barrier activity: Bar1 is an aspartyl protease with novel substrate specificity. MBio, 6(6), e01604–e01615. 10.1128/mBio.01604-15. [PubMed: 26604258]

Joshi S, Chen L, Winter MB, Lin Y-L, Yang Y, Shapovalova M, et al. (2017). The rational design of therapeutic peptides for aminopeptidase N using a substrate-based approach. Scientific Reports, 7(1), 1424. 10.1038/s41598-017-01542-5. [PubMed: 28465619]

Julien O, Zhuang M, Wiita AP, O'Donoghue AJ, Knudsen GM, Craik CS, et al. (2016). Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. Proceedings of the National Academy of Sciences of the United States of America, 113(14), E2001–E2010. 10.1073/pnas.1524900113. [PubMed: 27006500]

Kim H-R, Tagirasa R, & Yoo E (2021). Covalent small molecule immunomodulators targeting the protease active site. Journal of Medicinal Chemistry, 64(9), 5291–5322. 10.1021/acs.jmedchem.1c00172. [PubMed: 33904753]

Lentz CS, Ordonez AA, Kasperkiewicz P, La Greca F, O'Donoghue AJ, Schulze CJ, et al. (2016). Design of selective substrates and activity-based probes for hydrolase important for pathogenesis 1 (HIP1) from Mycobacterium tuberculosis. ACS Infectious Diseases, 2(11), 807–815. 10.1021/acsinfecdis.6b00092. [PubMed: 27739665]

Leontovyč A, Ulrychová L, O'Donoghue AJ, Vondrášek J, Marešová L, Hubálek M, et al. (2018). SmSP2: A serine protease secreted by the blood fluke pathogen Schistosoma mansoni with anti-hemostatic properties. PLoS Neglected Tropical Diseases, 12(4), e0006446. 10.1371/journal.pntd.0006446. [PubMed: 29677188]

Li H, Goh BN, Teh WK, Jiang Z, Goh JPZ, Goh A, et al. (2018). Skin commensal Malassezia globosa secreted protease attenuates Staphylococcus aureus biofilm formation. The Journal of Investigative Dermatology, 138(5), 1137–1145. 10.1016/j.jid.2017.11.034. [PubMed: 29246799]

Lohse MB, Brenes LR, Ziv N, Winter MB, Craik CS, & Johnson AD (2020). An opaque cell-specific expression program of secreted proteases and transporters allows cell-type cooperation in Candida albicans. Genetics, 216(2), 409–429. 10.1534/genetics.120.303613. [PubMed: 32839241]

Lysyk L, Brassard R, Arutyunova E, Siebert V, Jiang Z, Takyi E, et al. (2021). Insights into the catalytic properties of the mitochondrial rhomboid protease PARL. The Journal of Biological Chemistry, 296, 100383. 10.1016/j.jbc.2021.100383. [PubMed: 33556373]

Maffioli E, Jiang Z, Nonnis S, Negri A, Romeo V, Lietz CB, et al. (2020). High-resolution mass spectrometry-based approaches for the detection and quantification of peptidase activity in plasma. Molecules, 25(18), 4071. 10.3390/molecules25184071. [PubMed: 32899982]

Mahoney M, Damalanka VC, Tartell MA, Chung DH, Lourenço AL, Pwee D, et al. (2021). A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells. Proceedings of the National Academy of Sciences of the United States of America, 118(43), e2108728118. 10.1073/pnas.2108728118. [PubMed: 34635581]

Meyer NO, O'Donoghue AJ, Schulze-Gahmen U, Ravalin M, Moss SM, Winter MB, et al. (2017). Multiplex substrate profiling by mass spectrometry for kinases as a method for revealing quantitative substrate motifs. Analytical Chemistry, 89(8), 4550–4558. 10.1021/acs.analchem.6b05002. [PubMed: 28322550]

O'Donoghue AJ, Ivry SL, Chaudhury C, Hostetter DR, Hanahan D, & Craik CS (2016). Procathepsin E is highly abundant but minimally active in pancreatic ductal adenocarcinoma tumors. Biological Chemistry, 397(9), 871–881. 10.1515/hsz-2016-0138. [PubMed: 27149201]

O'Donoghue AJ, Knudsen GM, Beekman C, Perry JA, Johnson AD, DeRisi JL, et al. (2015). Destructin-1 is a collagen-degrading endopeptidase secreted by Pseudogymnoascus destructans, the causative agent of white-nose syndrome. Proceedings of the National Academy of Sciences of the Unites States of America, 112(24), 7478–7483. 10.1073/pnas.1507082112.

Rebello KM, McKerrow JH, Mota EM, O'Donoghue AJ, & Neves-Ferreira AGC (2018). Activity profiling of peptidases in Angiostrongylus costaricensis first-stage larvae and adult worms. PLOS Neglected Tropical Diseases, 12(10), e0006923. 10.1371/journal.pntd.0006923. [PubMed: 30379807]

Roncase EJ, Moon C, Chatterjee S, González-Páez GE, Craik CS, O'Donoghue AJ, et al. (2017). Substrate profiling and high resolution Co-complex crystal structure of a secreted C11 protease conserved across commensal bacteria. ACS Chemical Biology, 12(6), 1556–1565. 10.1021/acschembio.7b00143. [PubMed: 28414448]

Salcedo EC, Winter MB, Khuri N, Knudsen GM, Sali A, & Craik CS (2021). Global protease activity profiling identifies her2-driven proteolysis in breast cancer. ACS Chemical Biology, 16(4), 712–723. 10.1021/acschembio.0c01000. [PubMed: 33765766]

Sandholt GB, Stefansson B, Scheving R, & Gudmundsdottir Á (2019). Biochemical characterization of a native group III trypsin ZT from Atlantic cod (Gadus morhua). International Journal of Biological Macromolecules, 125, 847–855. 10.1016/j.ijbiomac.2018.12.099. [PubMed: 30550824]

Schneidman-Duhovny D, Khuri N, Dong GQ, Winter MB, Shifrut E, Friedman N, et al. (2018). Predicting CD4 T-cell epitopes based on antigen cleavage, MHCII presentation, and TCR recognition. PLoS One, 13(11), e0206654. 10.1371/journal.pone.0206654. [PubMed: 30399156]

Small JL, O'Donoghue AJ, Boritsch EC, Tsodikov OV, Knudsen GM, Vandal O, et al. (2013). Substrate specificity of MarP, a periplasmic protease required for resistance to acid and oxidative stress in Mycobacterium tuberculosis. The Journal of Biological Chemistry, 288(18), 12489–12499. 10.1074/jbc.M113.456541. [PubMed: 23504313]

Sojka D, Franta Z, Frantová H, Bartosová P, Horn M, Váchová J, et al. (2012). Characterization of gut-associated cathepsin D hemoglobinase from tick Ixodes ricinus (IrCD1). The Journal of Biological Chemistry, 287(25), 21152–21163. 10.1074/jbc.M112.347922. [PubMed: 22539347]

Sun J, LaRock DL, Skowronski EA, Kimmey JM, Olson J, Jiang Z, et al. (2020). The Pseudomonas aeruginosa protease LasB directly activates IL-1β. EBioMedicine, 60, 102984. 10.1016/j.ebiom.2020.102984. [PubMed: 32979835]

Winter MB, Greca FL, Arastu-Kapur S, Caiazza F, Cimermancic P, Buchholz TJ, et al. (2017). Immunoproteasome functions explained by divergence in cleavage specificity and regulation. ELife, 6, e27364. 10.7554/eLife.27364. [PubMed: 29182146]

Yoon MC, Ames J, Mosier C, Jiang Z, Podvin S, O'Donoghue AJ, et al. (2022). Distinct dibasic cleavage specificities of neuropeptide-producing cathepsin L and cathepsin V cysteine proteases compared to PC1/3 and PC2 serine proteases. ACS Chemical Neuroscience, 13(2), 245–256. 10.1021/acschemneuro.1c00653. [PubMed: 34986304]

Yoon MC, Christy MP, Phan VV, Gerwick WH, Hook G, O'Donoghue AJ, et al. (2022). Molecular features of CA-074 pH-dependent inhibition of cathepsin B. Biochemistry, 61(4), 228–238. 10.1021/acs.biochem.1c00684. [PubMed: 35119840]

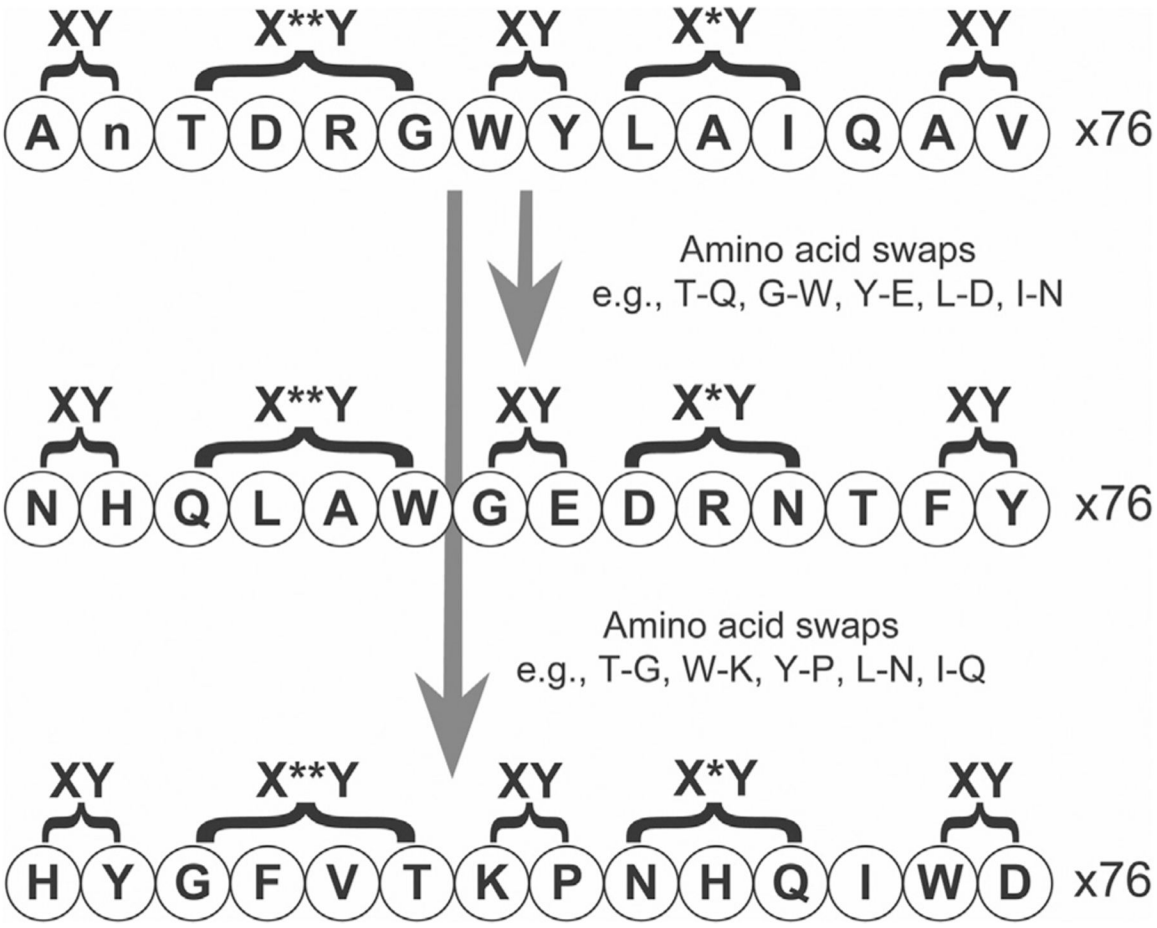

Fig. 1.

Rational design of MSP-MS library. A novel algorithm was developed to arrange amino acids pairs into a minimal number of decapeptides (76). Amino acids were then swapped with a physiochemically distinct counterpart twice, to afford two additional sets of 76 decapeptides. Pairs of amino acids were added to the termini to accommodate exopeptidase activity, affording a final library of 228 14-mers. n, norleucine.

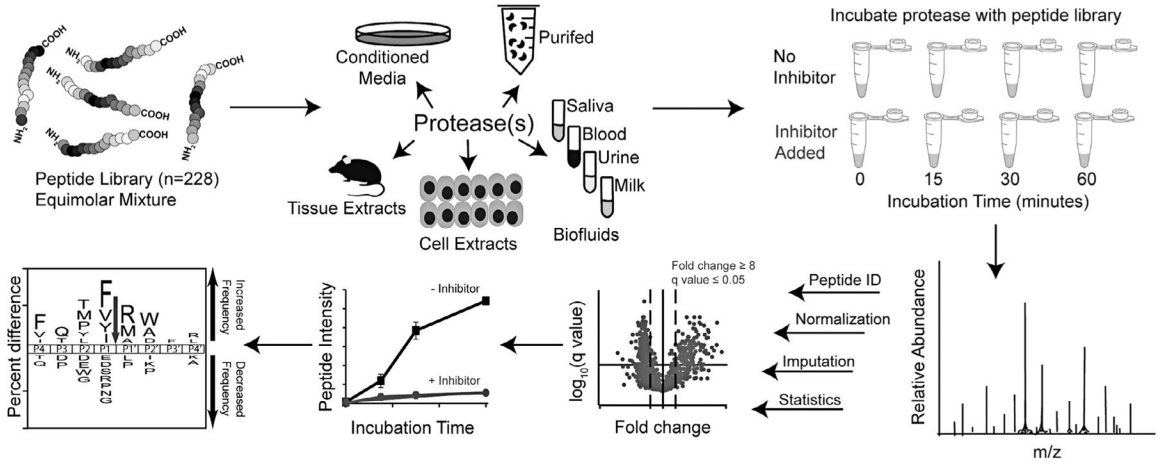

Fig. 2.

Workflow of MSP-MS experiment. A library of an equimolar mixture of 228 tetradecapeptides is incubated with a purified protease or a biological sample containing multiple proteases with or without a protease inhibitor. Aliquots of this assay are removed at multiple time intervals and the proteases are inactivated with a protein denaturant. Samples are prepared and subjected to LC-MS/MS for peptide identification and quantification. Cleavage products are identified by calculating the proteolytic kinetics from progression curves or relative abundance changes. The sequences of cleavage products are mapped back to library peptide sequences to identify the cleavage sites. The frequencies of amino acids identified from P4 to P4’ positions are calculated to generate specificity profiles.

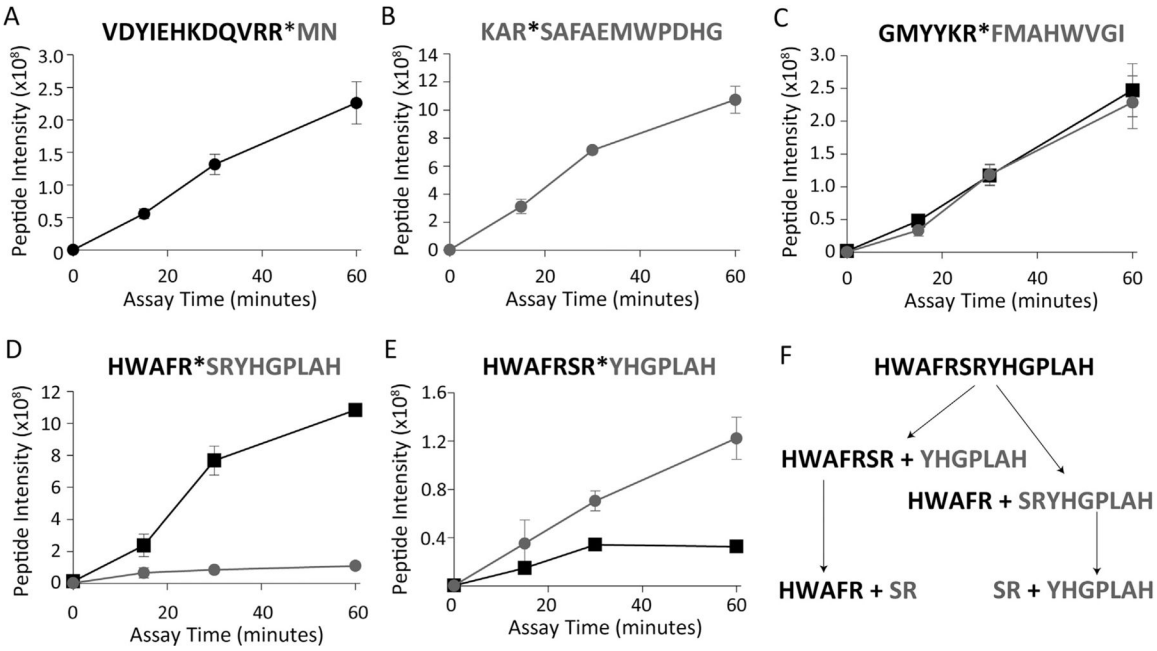

Fig. 3.

Progress curves for select peptide cleavage products within the 228-peptide library for hepsin MSP-MS experiment. Accumulation of (A) VDYIEHKDQVRR and (B) SAFAEMWPDHG over a 60-min timeframe. Smaller peptide fragments MN and KAR cannot be detected by MS/MS. (C) Accumulation of both cleavage products over time, where squares indicate the N-terminal product and circles indicate the C-terminal product. Hydrolysis of HWAFRSRYHGPLAH between (D) R and S and (E) R and Y. (F) The complete degradation pathway of HWAFRSRYHGPLAH by hepsin.

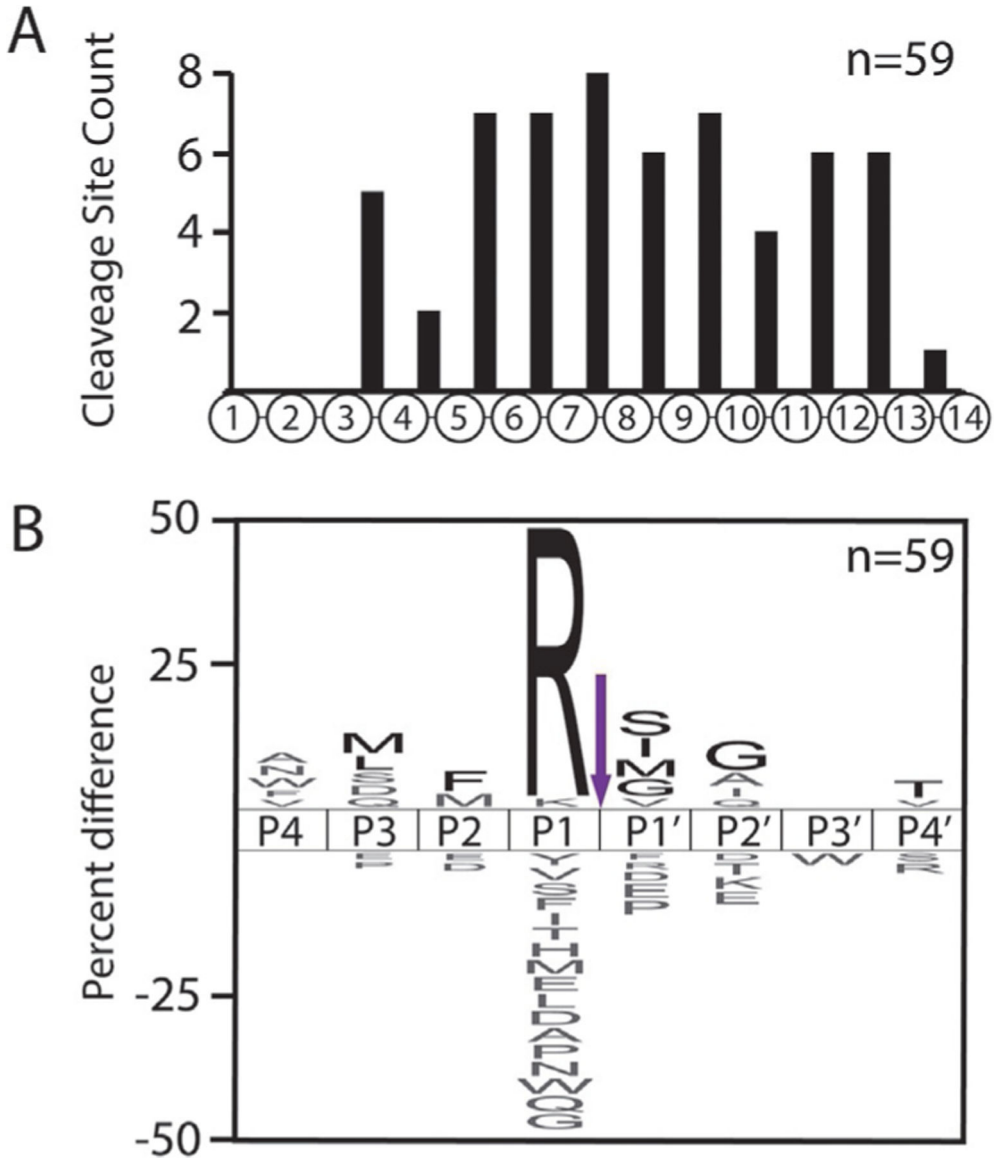

Fig. 4.

Specificity profile of hepsin as determined by MSP-MS. (A) Cleavage site distribution and substrates profile of human hepsin. Cleavages occur predominantly toward the center of the peptide, indicative of endopeptidase activity. (B) iceLogo representing substrate specificity of hepsin as derived from the MSP-MS experiment. M corresponds to norleucine, and bold residues in Panel B are significantly increased ($P < 0.05$).

Table 1

Design and sequences of core decapeptide library.

Parent Library	Switch	Sublibrary 1	Switch	Sublibrary 2
ATVYEFGHID	A and R V and K L and D I and N Q and T F and S W and G E and Y n and H P and P	RQKEYSWnNL	A and H V and R L and N I and Q T and G F and D W and K E and E n and S P and Y	HGRPEDTAQF
DKLnNWPQRR		LVDHIGPTAA		FWNSLKYIVV
RAVEGIKnWQ		ARKYWNVHGT		VHRETQWSKI
QSTYFHDLNP		TFQESnLDIP		InGPDAFNLY
NQAYGDnPSV		ITREWLHPFK		LIHPTFSYnR
VFILWRTEHH		KSNDGAQYnn		RDQNKVGEAA
HLPAEInQTT		nDPRYNHTQQ		ANYHEQSIGG
FDNRVGKWSY		SLIAKWVGFE		DFLVRTWKnP
SEDWAFKPTG		FYLGRSVPQW		nEFKHDWYGT
GLQVHnRYIN		WDTKnHAENI		TNIRASVPQL
IWTHNAGnSF		NGQnIRWHFS		QKGALHTSnD
FLREKQYDPV		SDAYVTELPK		DNVEWIPFYR
YYKRFnAHWV		EEVASHRnGK		PPWVDSHAKR
VDQELSGNTI		KLTYDFWIQN		RFIENnTLGQ
TDRGWYLAIQ		QLAWGEDRNT		GFVTKPNHQI
QFNVKSHPEE		TSIKVFnPYY		IDLRWnAYEE
ENYnVLTKAA		YIEHKDQVRR		ELPSRNGWHH
DSIRHQGPFW		LFNAnTWPSG		FnQVAITYDK
FPGQHRISDA		SPWTnANFLR		DYTIAVQnFH
KKTLVnYNEW		VVQDKHEIYG		WWGNRSPLEK
EPHSKVNFQQ		YPnFVKISTT		EYAnWRLDII
IALYWGRDTn		NRDEGWALQH		QHNPKTVFGS
NNGSLEQDVW		IIWFDYTLKG		LLTnNEIFRK
WHAnFRKYPI		GnRHSAVEPN		KAHSDVWPYQ
PDYQKERLFF		PLETVYADSS		YFPIWEVNDD
SnGANHTWVV		FHWRInQGKK		nSTHLAGKRR
IYRnHVQLGG		NEAHnKTDWW		QPVSARINTT
TPKFAWDESN		QPVSRGLYFI		GYWDHKFEnL
SWKGVRNDFT		FGVWKAILSQ		nKWTRVLFDG
TQnIEAPLHY		QTHNYRPDnE		GISQEHYNAP
HETRWLIFVS		nYQAGDNSKF		AEGVKNQDRn
SPnDGYAQNK		FPHLWERTIV		nYSFTPHILW
LLDHFYTSQI		DDLnSEQFTN		NNFADPGnIQ
nnKIGEVARP		HHVNWYKRAP		SSWQTERHVY
IQPWNnLKDD		ATPGIHDVLL		QIYKLSNWFF
IHGFEYVTAS		NnWSYEKQRF		QATDEPRGHn
GTAKNSSADI		WQRVIFFRLN		TGHWLnnHFQ

Parent Library	Switch	Sublibrary 1	Switch	Sublibrary 2
WnTFIVPPRS		GHQSNKPPAF		KSGDQRYYVn
WFSVIIPNLG		GSFKNNPIDW		KDnRQQYLNT
PYHKnETNIT		PEnVHYQINQ		YPAWSEGLQG
KHYSRQWWEn		VnEFATGGYH		WAPnVIKKES
SQTARWNDVD		FTQRAGILKL		nIGHVKLFRF
RSAFAEnWPD		AFRSRYHGPL		VnHDHESKYF
STSPHnIWAK		FQFPnHNGRV		nGnYASQKHW
QVFSWLNHYH		TKSFGDInEn		IRDnKNLAPA
NKRISQWnWE		IVANFTGHGY		LWVQnIKSKE
YQWQNSDPKA		ETGTIFLPVR		PIKILnFYWH
nRIYIEPETF		HANENYPYQS		SVQPQEYEGD
HTNDLPLISV		nQILDPDNFK		AGLFNYNQnR
LIQVPVKHEG		DNTKPKVnYW		NQIRYRWAET
PAPTYGRLWI		PRPQEWADGN		YHYGPTVNKQ
FNnYGYDLQT		SIHEWELDTQ		DLSPTPFNIG
RLNGEAVnFn		ADIWYRKHSH		VNLTEHRSDS
IEYKGKPTRW		NYEVWVPQAG		QEPWTWYGVK
GnLNIHFKFD		WHDINnSVSL		TSNLQADWDF
VYDWAFRIRS		KELGRSANAF		RPFKHDVQVn
DLHTVTAESL		LDnQKQRYFD		FNAGRGHEnN
AYKEQTNGNH		REVYTQIWIn		HPWEIGLTLA
GnGPFHIVKW		WHWPSnNKVG		TSTYDAQRWK
WQSPQVDLYD		GTFPTKLDEL		KInYIRFNPF
ESAEYNPTRL		YFRYEIPQAD		EnHEPLYGVN
QFTKGPKYEH		TSQVWPVEYn		IDGWTYWPEA
LGQRPTDNYE		DWTAPQLIEY		NTIVYGFLPE
RVYLTSPKAP		AKEDQFPVRP		VRPNGnYWHY
PHDVnGSRGL		PnLKHWFAWD		YAFRSTnVTN
DRWHTWnKIF		LAGnQGHVNS		FVKAGKSWQD
NDVNIFEIHG		ILKINSYNnW		LFRLQDEQAT
IEPDVSQVKH		NYPLKFTKVn		QEYFRnIRWA
KTnADSQESA		VQHRLFTYFR		WGSHFnIEnH
HNFHnAIGAR		nISnHRNWRA		ALDASHQTHV
AWKFGIFYLN		RGVSWNSEDI		HKWDTQDPNL
VLGWLNTSRQ		KDWGDIQFAT		RNTKNLGnVI
FKYIWYVQTA		SVENGEKTQR		DWPQKPRIGH
NnWSLYRnIR		IHGFDEAHNA		LSKnNPVSQV
nAFnKWHEGQ		HRSHVGnYWT		SHDSWKAETI
IWEPIDRGPW		NGYPNLAWPG		QKEYQFVTYK

Table 2

Full list of MSP-MS peptide library sequences.

LVATVYEFGHIDHM	HHFTQRAGILKLMP	MMVDQELSGNTIKV	KWLIHPTFSYMRWP	WFQHNPKTVFGSDL	PWKDMRQQYLNTGE
HIGLQVHMRYINVM	YHFQFPMHNGRVPD	EEPMLKHWFAWDQN	QTRDQNKVGEAAWY	YQLLTMNEIFRKWH	DLYPAWSEGLQGFW
KHPLETVYADSSEW	VMFLREKQYDPVSL	WILAGMQGHVNSTQ	REANYHEQSIGGWE	FLKAHSDVWPYQDA	DAWAPMVIKKESSI
MEFHWRIMQGKKAP	EVDNTKPKVMYWHF	QVILKINSYNMWAD	SMQFNVKSHPEELV	LEYFPIWEVNDDVK	PQMIGHVKLFRFNW
AQNEAHMKTDWWAY	EYPRPQEWADGNHL	VINYPLKFTKVMQQ	INDFLVRTWKMPGL	YWMSTHLAGKRRDW	WGVMHDHESKYFMQ
GDQPVSRGLYFITH	GFSIHEWELDTQKW	MIVQHRLFTYFRAW	RNMEFKHDWYGTPL	LWQPVSARINTTFD	WPMGMYASQKHWEE
YEQTHNYRPDMEWQ	EGADIWYRKHSHQL	TVMISMHRNWRAQH	IPTNIRASVPQLWW	YMDSIRHQGPFWML	WEIRDMKNLAPAND
TNMYQAGDNSKFQP	EDNYEVWVPQAGRN	TWRGVSWNSEDILT	IWQKGALHTSMDPQ	LDGYWDHKFEMLTW	DELWVQMIKSKEDQ
ENFPHLWERTIVNH	GMYYKRFMAHWVGI	LRKDWGDIQFATAN	IGDNVEWIPFYRDT	LNMKWTRVLFDGYQ	WNPIKILMFYWHEF
KQDDLMSEQFTNMY	ERWHDINMSVSLQI	NRSVENGEKTQRWI	IEPPWVDSHAKRNM	LQGISQEHYNAPDN	MLDKLMNWPQRRGM
TPHHVNWYKRAPNQ	EMKELGRSANAFQT	FRIHGFDEAHNAWM	VLRFIENMTLGQDY	LHAEGVKNQDRMGN	QMKKTLVMYNEWNL
PVIWTHNAGMSFMM	EPLDMQKQRYFDWL	AMTDRGWYLAIQAV	HYGFVTKPNHQIWD	LAMYSFTPHILWDE	WHSVQPQEYEGDNN
THATPGIHDVLLRP	EQREVYTQIWIMTV	TDHRSHVGMYWTLN	VFIDLRWMAYEEPW	LMNNFADPGMIQYL	WDAGLFNYNQMRGF
FPNMWSYEKQRFHP	GEWHWPSMNKVGQA	TQNGYPNLAWPGDI	QYELPSRNGWHHNP	MQSSWQTERHVYFA	DHNQIRYRWAETMW
PHWQRVIFFRLNTP	GNGTFPTKLDELQF	TEIWEPIDRGPWRF	RMENYMVLTKAAPV	MDQIYKLSNWFFEQ	WWYHYGPTVNKQLR
QEGHQSNKPPAFMH	GLYFRYEIPQADQW	LTHGRPEDTAQFWG	YAFMQVAITYDKWN	LLQATDEPRGHMYN	WMDLSPTPFNIGYK
DPGSFKNNPIDWHH	DITSQVWPVEYMQD	DTFWNSLKYIVVWT	DFDYTIAVQMFHEV	PETGHWLMMHFQED	DQVNLTEHRSDSEK
EHPEMVHYQINQKH	NVDWTAPQLIEYLD	RLVHRETQWSKING	FDWWGNRSPLEKMV	DMFPGQHRISDAEL	DNQEPWTWYGVKFR
RPVMEFATGGYHPN	WLAKEDQFPVRPTE	WTIMGPDAFNLYAL	FNEYAMWRLDIIPE	MWKSGDQRYYVMNE	PLTSNLQADWDFNR
NMRPFKHDVQVMLE	HLDWPQKPRIGHEM	AGSWKGVRNDFTEA	HTNKRISQWMWEIR	SKLGQRPTDNYEPS	WQKSNDGAQYMMGQ
DWFNAGRGHEMNLQ	FSLSKMNPVSQVLH	AVRAVEGIKMWQSM	MSYQWQNSDPKAGR	HRRVYLTSPKAPES	IQMDPRYNHTQQFP
PMEPHSKVNFQQHI	QHSHDSWKAETIEP	PGTQMIEAPLHYSA	ASMRIYIEPETFDK	WKPHDVMGSRGLYT	APSLIAKWVGFEPH
NGHPWEIGLTLALW	RTQKEYQFVTYKMD	QGHETRWLIFVSNA	KSHTNDLPLISVMR	PKDRWHTWMKIFNT	PDFYLGRSVPQWHE
NPTSTYDAQRWKNI	ADARKYWNVHGTHQ	RGSPMDGYAQNKHA	SLNQAYGDMPSVDM	LFNDVNIFEIHGVD	PYWDTKMHAENIAQ
NWKIMYIRFNPFLA	MGNNGSLEQDVWIA	GGLLDHFYTSQIPA	NTLIQVPVKHEGPK	SFIEPDVSQVKHME	LPNGQMIRWHFSEN
QLEMHEPLYGVNIE	AHLFNAMTWPSGHN	SAMMKIGEVARPLG	VSPAPTYGRLWIHK	GWKTMADSQESARD	VHSDAYVTELPKTN
NEIDGWTYWPEAVN	HDFGVWKAILSQPP	IARQPWNMLKDDMG	PSFNMYGYDLQTAK	KVHLPAEIMQTTQM	MPEEVASHRMGKNF
NDNTIVYGFLPEIW	HWAFRSRYHGPLAH	MAIHGFEYVTASRG	YTRLNGEAVMFMSK	EWHNFHMAIGARSD	GPKLTYDFWIQNLP

LVATVYEFGHIDHM	HHFTQRAGILKLMP	MMVDQELSGNTIKV	KWLIHPTFSYMRWP	WFQHNPKTVFGSDL	PWKDMRQQYLNTGE
QAVRPNGMYWHYLM	HPTKSFGDIMEMLY	WAGTAKNSSADIQG	ESIEYKGKPTRWQR	FYAWKFGIFYLNGD	NLSEDWAFKPTGPM
QWYAFRSTMVTNIY	ADIVANFTGHGYHQ	EAWMTFIVPPRSAG	SSGMLNIHFKFDWR	KYVLGWLNTSRQAE	NHQLAWGEDRNTFY
NNFVKAGKSWQDIP	ALETGTIFLPVRHD	HAWFSVIIPNLGDG	IKVYDWAFRIRSGT	MYFKYIWYVQTADD	AETSIKVFMPYYGH
HNLFRLQDEQATVG	AWHANENYPYQSKL	GIQSTYFHDLNPYM	GRDLHTVTAESLMS	AYNMWSLYRMIRQE	VDYIEHKDQVRRMN
HMIALYWGRDTMFI	ANMQILDPDNFKRE	AAPYHKMETNITSG	MRAYKEQTNGNHVS	NFMAFMKWHEGQYE	MHSPWTMANFLRGP
HEQEYFRMIRWAVL	LGWHAMFRKYPIMA	NAKHYSRQWWEMFG	AKGMGPFHIVKWAS	DDRQKEYSWMNLIQ	HQVVQDKHEIYGDP
QQWGSHFMIEMHVI	SGPDYQKERLFFWA	PASQTARWNDVDGG	FIVFILWRTEHHAM	RDLVDHIGPTAAYH	GQYPMFVKISTTHW
QNALDASHQTHVLL	FGSMGANHTWVVKA	KARSAFAEMWPDHG	RKWQSPQVDLYDKS	GHTFQESMLDIPKQ	PPNRDEGWALQHTF
VYHKWDTQDPNLDH	DGIYRMHVQLGGAA	GASTSPHMIWAKPG	QKESAEYNPTRLHS	ELFDNRVGKWSYRM	NQIIWFDYTLKGEH
QDRNTKNLGMVIEG	HGTPKFAWDESNGA	GSQVFSWLNHYHRK	DKQFTKGPKYEHSS	PNITREWLHPFKVH	QPGMRHSAVEPNGW

M is norleucine.