Natural Product Reports

Leveraging orthogonal mass spectrometry based strategies for comprehensive sequencing and characterization of ribosomal antimicrobial peptides natural products

Journal:	*Natural Product Reports*
Manuscript ID	NP-REV-07-2020-000046.R1
Article Type:	Review Article
Date Submitted by the Author:	28-Aug-2020
Complete List of Authors:	Moyer, Tessa; University of North Carolina at Chapel Hill, Chemistry Parsley, Nicole; University of North Carolina at Chapel Hill, Chemistry Sadecki, Patric; University of North Carolina at Chapel Hill, Chemistry Schug, Wyatt; University of North Carolina at Chapel Hill, Chemistry Hicks, Leslie; University of North Carolina at Chapel Hill, Chemistry

Leveraging orthogonal mass spectrometry based strategies for comprehensive sequencing and characterization of ribosomal antimicrobial peptides natural products

Tessa B. Moyer, Nicole C. Parsley, Patric W. Sadecki, Wyatt J. Schug, and Leslie M. Hicks

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Ribosomal antimicrobial peptide (AMP) natural products, also known as ribosomally synthesized and post-translationally modified peptides (RiPPs) or host defense peptides, demonstrate potent bioactivities and impressive complexity that complicate molecular and biological characterization. Tandem mass spectrometry (MS) has rapidly accelerated bioactive peptide sequencing efforts, yet standard workflows insufficiently address intrinsic AMP diversity. Herein, orthogonal approaches to accelerate comprehensive and accurate molecular characterization without the need for prior isolation are reviewed. Chemical derivatization, proteolysis (enzymatic and chemical cleavage), multistage MS fragmentation, and separation (liquid chromatography and ion mobility) strategies can provide complementary amino acid composition and post-translational modification data to constrain sequence solutions. Examination of two complex case studies, gomesin and styelin D, highlight the practical implementation of the proposed approaches. Finally, we emphasize the importance of heterogeneous AMP peptidoforms that confer varying biological function, an area that warrants significant further development.

Outline

1. Introduction
2. Amino acid composition
    - 2.1. Cysteine
    - 2.2. Methionine
    - 2.3. Charged residues
    - 2.4. Leucine and Isoleucine
    - 2.5. Asparagine
3. Post-translational modifications
    - 3.1. Terminal modifications
    - 3.2. Dehydration
    - 3.3. Cyclization
    - 3.4. Oxidation/hydroxylation
    - 3.5. Glycosylation
    - 3.6. Halogenation
    - 3.7. Stereoisomers
4. Case studies (styelin D and gomesin)
5. Peptidoform heterogeneity and functional implications
6. Conclusions

1. Introduction

Ribosomally synthesized antimicrobial peptides (AMPs) are a prominent class of small (<10 kDa) host defense peptides ubiquitously identified across all domains of life.[1] In unicellular organisms they are involved in niche protection, while in complex organisms AMPs are part of the innate immune system.[2,3] Although AMPs were originally recognized for their antibacterial activity, they demonstrate a wide range of biological targets[4,5] and activities.[6,7] Generally cationic and hydrophobic, AMPs represent a diverse range of sequences with highly variable post-translational modifications (PTMs),[1,8] though fully processed AMPs adopt a range of conserved secondary structures often used to classify related peptides.[9,10] Heterogenous populations with varying PTM localization or occupancy adds an additional level of complexity.[11–15] Novel AMP discovery is driven predominately by bioassay-guided fractionation and genome mining,[16–20] while repositories such as the Antimicrobial Peptide Database (APD), PhytAMP, and Cybase provide centralization for sequences and facilitate comparisons of sequence diversity and homology.[8,21–25]

Mass spectrometry (MS) has become the dominant analytical technology leveraged for sequence characterization of AMPs. Unlike Edman degradation,[26,27] MS-based strategies have minimal sample requirements and can analyze peptides in mixtures and/or with blocked N-termini.[28] Genome mining based methods can be leveraged to identify likely peptide sequences but require a species specific database and knowledge of the expected biosynthetic processing to predict PTMs.[29] MS-based methods have been developed that do not rely on genomic information and enable direct measurement of PTMs. High resolving power mass analyzers provide accurate intact mass measurements and complementary tandem mass spectrometry (MS/MS) methods reveal primary sequence and PTM localization.[30,31] Integrated genomics, transcriptomics, and bioinformatics approaches are increasingly used to facilitate database searching,[32,33] *de novo* sequencing,[34–36] and molecular networking,[29,37] which have been well reviewed elsewhere.[29,32–36] Even so, the combinatorial explosion of sequence solutions from large biomolecules with diverse modifications can result in inaccurate and/or incomplete characterization.[29] Herein, a compendium of tractable orthogonal approaches, broadly categorized in four areas (Figure 1), that complement MS/MS-based sequencing are reviewed to enable accelerated, accurate AMP molecular characterization.

Chemical derivatization of select functional groups results in predictable mass shifts that can identify the presence and stoichiometry of specific amino acids, as well as the modification status of termini (Figure 1A, Table 1).[38–43] This class of methods is most effective for sequence features that have unique and chemically reactive functional groups.

Enzymatic/chemical digestion can reveal specific amino acid residues and post-translational modifications based on cleavage specificity (Figure 1B, Table 2).[44–53] Digestions can be used to produce smaller peptides with less complex fragmentation spectra than the intact AMP, though reconstruction of the undigested, intact sequence is required - making this most effective for relatively pure AMPs. Although digestions are not perfectly efficient or specific, they yield information about sequence features that lack chemically reactive functional groups and enable differentiation between those that share identical functional groups.

Multistage MS-based strategies are information-rich, consume minimal sample, and can produce diagnostic fragment ions for a variety of sequence features (Figure 1C, Table 3).[54–62] Complementary MS/MS fragmentation methods (e.g. CID, HCD, EThcD, UVPD) produce different fragmentation patterns, stability of labile PTMs, and sequence information.[31]

Separations (Figure 1D) can reveal retention differences that can be essential for resolving modifications that do not alter peptide mass (e.g. stereoisomers) but do alter other physicochemical properties such as hydrophobicity or conformation.[63] The focus herein is for online separations applied to analysis of complex extracts, as opposed to those used for peptide isolation.

Essential sequence information including amino acid composition (Section 2) and possible modifications (Section 3) constrains possible sequence solutions and accelerates accurate molecular characterization while simultaneously minimizing the necessity for peptide isolation. Practical implementation of the included approaches is demonstrated via two complex case studies (Section 4). Finally, elaborated AMP heterogeneity and corresponding critical functional implications related to the biological impact are explored (Section 5).

2. Amino acid composition

AMPs represent highly diverse primary sequences to support their various bioactivities.[8] The frequency of amino acids in antimicrobial peptides differs from the Swiss-Prot database of annotated proteins (Figure 2A).[8,64] Most notably, AMPs are enriched in cysteine, arginine, and lysine, reflecting the importance of disulfide bonds and cationic properties to AMP structure and function.

The primary source of sequence information during MS-based peptide sequencing are backbone fragments (e.g. *b*-, *y*- ions),[31] yet other product ions (e.g. immonium / neutral loss) can support the presence of specific amino acid residues (Table 3).[54–57] Product ion generation is influenced by residue position, structure, fragmentation type, and experimental conditions.[54,65,66] As such, residue-specific fragment ions can provide a starting point for the analysis of amino acid composition, but cannot exclude residues from sequences or quantify the number of a specific residue. Free online tools (Protein Prospector's MS-Product, www.prospector.ucsf.edu or MS/MS Fragment Ion Calculator, http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html) predict fragmentation of user defined sequences and PTMs, generating theoretical fragments lists usefully for quickly assessing possible sequences.

Tandem mass spectrometry of an intact AMP is often insufficient for full sequence characterization due to AMP length and complexity. Complementary experiments facilitate imposition of sequence constraints to improve residue assignment accuracy and differentiate isomeric (Ile/Leu) / isobaric (Gln/Lys) residues. Derivatization strategy implementation depends on side-chain chemical reactivity and functional group uniqueness.[67]

Herein, approaches that can be applied to sulfur-containing (Cys, Met), basic (Lys, Arg), acidic (Glu, Asp), isomeric (Leu, Ile), and the polar residue Asn are assessed. These residues are critical to AMP structure/function or their characterization presents significant analytical challenges.

2.1. Cysteine (Cys)

Cysteine (Cys, pI 5.07) is a thiol-containing amino acid whose occurrence generally increases with organismal complexity (from 0.5% in Archaea to 2.6% in mammals)[68] and is prominently represented in AMPs (Figure 2A).[8] Cys residues can form intramolecular disulfide bonds often critical for structure/activity and can be diagnostic of specific AMP families.[1] The presence of disulfide bonds can reduce MS/MS fragmentation efficiency, thus preventing sequencing of the intact peptide. Cysteine derivatization can define AMP Cys content, constrain sequence space, and improve peptide fragmentation.

Chemical derivatization – Alkylation: Cysteine alkylation is used extensively in proteomics applications and has been frequently applied for AMP characterization (Table 1).[69,70] Peptides are first chemically reduced to the free thiol form before alkylating agents covalently add a defined moiety. A variety of alkylating agents are commercially available, with iodoacetamide (IAM), methyl methanethiosulfonate (MMTS), and N-ethylmaleimide (NEM) being among the most common (Table 1).[38,39] MS analysis of samples before and after alkylation reveal mass shifts dependent on alkylating agent mass and the corresponding the number of Cys residues present in the sequence.

Critically, this method can be used to differentiate between Cys involved in disulfide bonds (oxidized) and those in free thiol form (reduced) as disulfide-bound Cys produce a mass shift 1 Da greater than free Cys (due to lack of hydrogen atoms in S-S). Also, mass shifts corresponding disulfide bonds must occur in pairs because two Cys participate in each disulfide bond. The crucial role of Cys to AMP structure and the accessibility of alkylating agents makes this an essential component of the AMP characterization toolbox.

2.2. Methionine (Met)

Methionine (Met, pI 5.74) is the N- terminal residue for most eukaryotic proteins and is often cleaved from mature AMPs.[71] The overall frequency of Met is decreased in AMPs relative to the annotated proteome (Figure 2A).[8,64] Met supports the formation of α-helices impacting peptide secondary structure.[72,73] This sulfur-containing residue is readily oxidized during sample preparation and precautions (e.g. buffer and temperature) must be taken to control biologically relevant redox states.[74]

Chemical derivatization – Alkylation: Purposeful Met alkylation is rare in proteomics and is generally observed as a byproduct of cysteine alkylation.[39,75] Despite this, Met alkylation with common alkyl halides (e.g. IAM) could easily be implemented for AMP characterization because Met is the only nucleophilic residue that remains deprotonated under acidic conditions, allowing pH-controlled selective derivatization (Table 1).[40,76,77] Met residues derivatized with IAM produce diagnostic fragment ions (Table 3) that confirm residue identity.[59]

Chemical cleavage - Cyanogen Bromide: Met residues can be detected via cleavage with cyanogen bromide (CNBr), a reaction extensively used in Edman degradation and MS-based sequencing of AMPs.[78–81] Incubation with CNBr under acidic conditions results in the conversion of methionine to homoserine, followed by C-terminal amide bond cleavage (Table 2). Cleavage efficiency is reduced at Met-Ser and Met-Thr bonds as homoserine-Thr and homoserine-Ser are stable without bond cleavage.[82] Specificity is tenuous, as CNBr has been shown to be equally selective to Met and Tyr residues, may cleave at aspartic and glutamic acid, and produces undesirable side products resulting in the loss of methionine side-chains.[83] Additional byproducts formed via the oxidation of methionine to methionine sulfoxide by CNBr can be reduced using 70% formic acid reaction conditions.[44] As a note, CNBr is considered highly hazardous to human health through all exposure routes including causing pulmonary edemas if inhaled.[84] Thus, Met alkylation is a more attractive strategy than CNBr cleavage given its lack of specificity, high rate of side product formation, and known safety concerns.

2.3. Charged residues (Lys, Arg, Glu, and Asp)

Generally, AMPs are composed of more basic residues and fewer acidic residues than a standard proteome (Figure 2A).[8,64] Containing an abundance of lysine (Lys, pI 9.74) and arginine (Arg, pI 10.76),

AMPs are generally positively charged at physiological pH. Even so, anionic AMPs containing many glutamic acid (Glu, pI 3.22) and aspartic acid (Asp, pI 2.77) residues have been identified in vertebrates, invertebrates, and plants.[85,86] Understanding the charged residue composition of an unknown AMP has a myriad of benefits - potentially elucidating peptide class, specific sequence features, and isobaric differentiation (Lys/Gln). Defining net charge is also useful for developing ion exchange-based chromatographic methods for peptide fractionation/isolation. Complementary chemical derivatizations and/or enzymatic cleavages can be used in combination to explore AMP charged residue composition.

Chemical derivatization – Dimethylation (Lys) and methyl esterification (Glu and Asp): Lysine residues can be identified and quantified using dimethyl labeling, which modifies primary amine functional groups within the peptide (i.e. Lys and N-termini) (Table 1).[41,87–89] Methyl esterification is a common approach that converts carboxylic acid (-COOH) functional groups (i.e. Glu, Asp, and C-termini) to their corresponding methyl ester (-$COOCH_3$) (Table 1).[42,90–92] Care must be taken in reaction conditions to prevent side products such as over methylation to the quaternary salt or formation of N-methyl-4-imidazolidinone.[89,93,94] Both approaches produce predictable mass shifts based on the number of derivatized sites; however, to unambiguously enumerate Lys or acidic residues, one must determine if the peptide termini are contributing to the observed mass shift.[51,95] PTMs such as pyroglutamic acid, acetylation, amidation and peptide cyclization can block termini from derivatization. Thus, methods to identify terminal modifications (Figure 4, discussed in detail in Sections 3.1/3) can clarify charged residue content.

Notably, dimethylation can be applied in a more focused manner to distinguish isobaric Lys and Gln (Δm= 0.0434 Da) in low resolving power MS/MS spectra.[34,96] When an AMP sequence is known with the exception of assigning ambiguous Lys/Gln, dimethylation mass shift can define the number of Lys. The number of Gln can be deduced based on the total Lys/Gln sites and known number of Lys. Similar mass shift analysis of fragment ions can be used to determine the position of each Lys or Gln.

Enzymatic cleavage. Lys and Arg can be identified and differentiated based on protease specificity (Trypsin, Lys-N, Lys-C, Arg-C, Table 2). Trypsin is a serine protease that hydrolyzes the C-terminal amide bond of Lys and Arg residues, unless followed by proline, with high specificity.[45] Tryptic digestions can be used to confirm that AMPs contain basic residues. Additional digestions using proteases with higher specificity (Lys-N, Lys-C, Arg-C) can be used to detect the presence of specific residues.[97] While chemical derivatization can enumerate the total number of acidic residues, endoproteinase digestion (Glu-C, Asp-N, Table 2) can differentiate Glu and Asp. Glu-C is a serine protease with cleavage specificity for Glu and Asp.[45] In ammonium bicarbonate/acetate buffering systems, Glu-C cleaves preferentially at the C-terminus of Glu;[46] however, it loses specificity and cleaves at both Asp and Glu residues in phosphate buffered systems. Alternatively, endoproteinase Asp-N can be used to hydrolyze N-terminal to aspartic acid residues.[45,98]

2.4. *Leucine/Isoleucine (Leu/Ile)*

The branched-chain amino acids leucine (Leu) and isoleucine (Ile) are structural isomers with identical exact mass (internal residue monoisotopic mass = 113.0841 Da). Leu is less common in AMPs but the frequency of Ile remains similar to the overall proteome (Figure 2A).[8,64] Leu/Ile residues have significant biological consequences, affecting peptide activity, binding, and expression.[99,100] Although Leu/Ile are the most challenging residues to assign via MS, they are readily distinguished using DNA/RNA methods, Edman degradation, and multidimensional NMR.[101–103] These methods require databases of genetic information or sufficient quantities of purified AMP which may not be readily available for all

peptides.[101–103] Here, proteolytic and multistage mass spectrometry methods to rapidly differentiate Ile/Leu isomers, without the use of genetic information or purified peptides are highlighted.

Enzymatic cleavage – Chymotrypsin: Chymotrypsin is an endoproteinase that cleaves at Leu but not Ile (Table 2).[45] This preferential cleavage produces peptides containing N-terminal Leu and subsequent MS/MS sequencing can confirm the identity of cleavage products.[70] However, a lack of observed cleavage cannot definitively assign Ile residues. Missed cleavages at Leu-positions due to N-terminal proline or low abundance products can occur, and caution must be exercised to avoid inaccurate interpretation.

Multistage MS: MS^n of Leu/Ile-containing peptides produces characteristic side-chain losses that differentiate Leu/Ile residues and are detected via multistage fragmentation of *z*-ions or immonium ions.[58,104–106] These methods have been successfully applied to differentiate Leu/Ile in AMPs without prior DNA/Edman sequencing data.[107,108]

Fragmentation of *z*-ions ending in Leu or Ile produce diagnostic side-chain losses, 43.0548 Da and 29.0391 Da, respectively.[104,106] This method relies on the formation of *z*-ions with Leu/Ile at the N-terminus.[104] Eighty-one percent of Leu/Ile sites from a set of non-tryptic disulfide bound peptides from *Rana ridibunda* were successfully assigned using *z*-ion fragmentation illustrating the potential impact of this method on cysteine rich AMP sequence characterization.[103]

Multistage fragmentation of Leu/Ile immonium ions is also an effective discrimination strategy (Figure 3).[58,104] MS^2 of Leu/Ile containing peptides produce identical 86.0970 *m/z* immonium ions (Figure 3B), and subsequent MS^3 analysis of the immonium ion generates a diagnostic 69.0578 *m/z* ion produced in high abundance from Ile residues.[58,104,109] For Leu, the 69.0578 *m/z* ion is <10% of the precursor abundance (Figure 3C). This method is useful for peptides containing only a single Leu/Ile. Peptides containing multiple Leu/Ile require additional fragmentation steps to correctly assign each Leu/Ile position because a fragment ion containing a single Leu/Ile (Figure 3D and E) must first be produced.[104] Alternatively, enzymatic digestions that cleave the AMP into peptides that only contain a single Leu or Ile can be coupled with MS^n to limit the number of required fragmentation stages. Enzymes that cleave both Leu and Ile, e.g. thermolysin, can be used to guarantee only a single Leu/Ile in each enzymatically-digested fragment, assuming no missed cleavages.[47,109]

2.5. *Asparagine (Asn)*

Asparagine (Asn) is a polar amino acid with a carboxamide side-chain and the same exact mass as a Gly-Gly diamino acid (114.0429 Da). Poor AMP fragmentation (i.e. no cleavage between Gly-Gly) may result in an incorrect sequence assignment. Practical chemical derivatization of the non-ionizable, polar Asn remains a largely unanswered challenge as the most developed strategy requires dirhodium metallopepide catalysts and alkylates both asparagine and glutatmine.[67,110] Chemical cleavage can be used to identify the presence of Asn and confirm sequencing assignments.

Chemical cleavage - Diacetoxyiodobenzene: Peptides can be cleaved N-terminal to Asn using diacetoxyiodobenzene (Table 2).[48,111] This reaction includes the Hoffman rearrangement of the Asn side-chain, cyclization between the Asn side-chain and N-terminus, and cleavage N-terminal to Asn residues.[48] Peptide N-termini must be blocked (e.g. N-Fmoc-protected, dimethylation) or cleavage N-terminal to Asn resides will not occur.[48] In cases where an AMP sequence is known but Asn/Gly-Gly assignment is ambiguous, detection of diacetoxyiodobenzene cleavage products can confirm Asn.

2.6 Summary

AMPs are a diverse class of peptides with unique features, as reflected by differential amino acid frequencies (Figure 2A). Here, we have highlighted strategies to identify or enumerate nine amino acid residues (Cys, Met, Lys, Arg, Glu, Asp, Leu, Ile, Asn) essential in forming important structural features, such as disulfide bonds and nonpolar faces, that contribute to the efficacy of the peptide. Accurate assignment of isobaric residues or diamino acids can be imperative for recapitulation of biological activity[99,100] For example, aurein 2.2 and 2.3 vary only in a single Leu to Ile mutation yet 2.2 causes greater membrane leakage in *Staphylococcus aureus*.[112] Experiments to define amino acid composition can be used to confine possible sequence solutions and differentiate residues of the same mass, thus resulting in more efficient and accurate AMP characterization.

3. Post-translational Modifications

Antimicrobial peptides contain a diverse array of post-translational modifications (PTMs) that modulate chemical properties, biological activity, and stability (Figure 2B). PTMs increase AMP sequence diversity and are not easily predicted from genomic data, making molecular characterization challenging. Although commonly identified modifications can be considered during sequencing, natural AMP diversity demonstrates that unusual modifications need be considered for comprehensive characterization. Targeted strategies to identify PTMs are elaborated herein with a focus on practical methods which address experimental uncertainties. For example, a single method may not be sufficient to differentiate between terminal modifications and series of experiments may be required for clarification (Figure 4). PTMs which impact peptide stereochemistry but not mass are often missed by standard workflows and may require extremely targeted experiments. Together, these methods can constrain possible sequence assignment and ensure that appropriate PTMS are considered.

3.1 Terminal modifications

Peptide termini modification can enhance AMP resistance to proteolysis, increase the half-life of AMPs, and may be critical for bioactivity.[113,114] C-terminal amidation is the highest frequency PTM found on peptides deposited in the APD (Figure 2B).[8,115,116] Pyroglutamic acid and acetylation of the N-terminus are less common.[8,92,117] While varying modifications occur on the respective termini, strategies to detect and, where possible, characterize modifications are similar (Figure 5).

Detecting blocked termini. Chemical derivatization and enzymatic cleavage methods are available to reveal blocked termini. When a PTM is present, the terminus is blocked from modification and no mass shift is observed.[92,118,119] Common derivatization schemes include dimethyl labeling of the N-terminal primary amine[95] (Section 2.3, Table 1) and methyl esterification of the C-terminal carboxyl group[42] (Section 2.3, Table 1). However, application of this approach is challenged by a lack of specificity for termini, where charged residues sharing identical functional groups (e.g. lysine and N-terminus, acidic residues and the C-terminus) are also derivatized and convolute the resulting analysis. Thus, enumeration of these residues is required to accurately predict expected mass shifts and confirm status of termini modification.

In cases where the number of Lys, Glu, or Asp is unknown, it may be preferable to identify terminal modifications via digestion with non-specific exopeptidases (i.e. carboxypeptidases or aminopeptidases) (Table 2). These enzymes iteratively remove residues from the N- or C-terminus eliminating unmodified peptides and leaving modified peptides intact.[49,69,120] There are some scenarios when exopeptidases can provide false positives because they will not cleave for reasons other than terminal modifications (e.g. D-amino acids or proximity of proline).[50]

Differentiating termini modifications. Some terminal modifications can be specifically identified using MS/MS fragmentation patterns or enzymatic digestions (Figure 5). Amidated peptides exhibit a prominent ammonia neutral loss from the protonated precursor following fragmentation (Table 3).[60] This is a particularly appealing strategy because data may be collected during standard LC-MS/MS profiling. However, peptides with C-terminal Asp or Glu residues produce less intense ammonium losses and are a potential source of false positives.[60] N-termini can be differentiated using modification specific aminopeptidases (e.g. pyroglutamate aminopeptidase or acyl-amino acid releasing enzymes) which remove modified residues from the N-terminus (Table 2).[51,121,122] These highly specific peptidases are used extensively to facilitate Edman degradation and are easily adapted to an MS workflow.[69,121–124]

3.2 Dehydration

Serine (Ser) and threonine (Thr) can be enzymatically dehydrated to form dehydroalanine (Dha) or dehydrobutyrine (Dhb), a common transformation that is particularly relevant to lanthipeptides and cyanobactins (Figure 2B).[1,8] Dehydration modifications can impact bioactivity such as a critical role in nisin-lipid II binding affinity.[125] Dha and Dhb can be further modified to form thioether bridges (see Section 3.3).[126] Dehydrated residues can be chemically derivatized via reductive desulfurization (Table 1) which targets the side-chain alkene bond forming Ala from Dha or α-aminobutyric acid (Abu) from Dhb via the addition of two protons.[43,127] This reaction also modifies dehydrated residues involved in thioether bridges (Section 3.3) producing two Ala or Ala and Abu.[127] Although, reductive desulfurization results in a different mass shift for free Dha/Dhb (+2 H) and those involved in a thioether bond (+1H), the derivatized products are identical hindering localization when dehydrated residues and thioether bridges occur on the same peptide. However, deuterated reactions can facilitate discrimination by taking advantage of the different number of deuterium atoms added in place of hydrogen - producing Ala/Abu which are differentially deuterated based on their participation in thioether bridges.[43] The now different mass residues can be readily distinguished during *de novo* sequencing.

3.3 Cyclization

Peptide cyclization accounts for four of the most common APD modifications (i.e. disulfide bonds, backbone cyclization, thioether bridge, and side-chain to backbone cyclization) (Figure 2B).[8] Cyclization reduces conformational entropy and susceptibility to degradation by peptidases. Peptides can be cyclized with a normal peptide bond between the N- and C-termini ("head-to-tail") or different combinations of side-chain/side-chain or side-chain/terminus connections. Identification of cyclization is often the first step in AMP characterization, where the type is often highly conserved among related AMPs [e.g. cyclotides (head-to-tail cyclized),[128] lanthipeptides (thioether bridges),[129] and lasso peptides (side-chain/N-terminus cyclization)[130]]. In many cases, cyclization reduces MS/MS fragmentation efficiency and peptides must be linearized to obtain sufficient fragmentation for sequence elucidation.[131] Here, we discuss methods to identify cyclization and support sequencing of cyclized peptides.

Identification. Identifying cyclization can be challenging given the breadth of possible connectivity. Here, cyclizations are grouped by those that involve peptide termini (side-chain/termini and head-to-tail) and those that do not (side-chain/side-chain). Cyclization that involves the termini can be detected using similar strategies to terminal modifications (Section 3.1). Different types of cyclization involving termini produce varying -COOH and -NH_2 derivatization results which can be used as a first step to characterize AMP cyclization (Figure 5). For example, neither termini of head-to-tail cyclized peptides

can be derivatized, where a side-chain/termini cyclized peptide will have one terminus available for derivatization. However, this approach does not provide definitive evidence regarding cyclization alone.

Exopeptidase incubation can provide additional evidence for cyclization - as cyclic peptides are resistant to proteolytic degradation and remain intact after prolonged incubation. They will also be resistant to more specific peptidases like pyroglutamate aminopeptidase or acyl-amino acid releasing enzymes. Again, this only provides indirect evidence of cyclization involving peptide termini and indicated that additional steps may be needed to enhance MS/MS fragmentation for peptide sequencing.

Side-chain/side-chain cyclization requires targeted strategies. Disulfide bonds are the most common type and can be detected using cysteine alkylation (Section 3.1, Table 1). Cysteines involved in disulfide bonds produce a characteristic mass shift after reduction / alkylation that is dependent on the alkylating agent. AMPs within a specific class often share the same number of disulfide bonds. For example, AMPs belonging to the cyclotide family can be predicted from a reduction/alkylation mass shift of +348.1756 Da, consistent with three disulfide bonds / six cysteine residues modified with iodoacetamide.[70] Thioether bridges are another type of side-chain/side-chain cyclization and are formed by covalent bonds between a Cys thiol and a Ser or Thr.[127] Reductive desulfurization can be used to identify thioether bonds (Section 3.3, Table 1) and deuterated reaction conditions discriminate between thioether bridges and dehydrated residues.[43,127]

Often peptides have several possible side-chain/side-chain and side-chain/termini connections and correct linkages cannot be predicted based on sequence alone. Strategic intact endoproteinase digestions (e.g. Glu-C, chymotrypsin, pepsin) can produce cross-linked digest products which reveal internal connectivity. This method is most commonly used to determine Cys-Cys linkages within AMPs. Microwave-assisted partial acid hydrolysis can be used to cleave the peptide backbone in AMPs resistant to proteolytic degradation.[132]

Sequence elucidation of cyclic species. Cyclized peptides present unique challenges when using MS-based methods to elucidate sequence. MS^2 experiments with intact cyclic peptides often result in linearization via a single cleavage event, yielding no sequence information.[131] A second fragmentation event is required to generate sequence information. This is particularly problematic in head-to-tail cyclized peptides whose initial linearization can occur anywhere along the peptide backbone, thus complicating the spectrum with numerous, redundant fragment ions.[131] Site-specific intentional linearization strategies can be implemented to enhance and simplify fragmentation of cyclized peptides.

Chemical derivatizations and enzymatic cleavages can be used to linearize peptides prior to MS. Chemical derivatizations target specific side-chain/side-chain connections, such as disulfide bonds or thioether bridges, to produce linear peptides.[38,127] Strategic endopeptidase cleavages can be used to linearize head-to-tail or side-chain/terminus cyclizations.[70,131] However, sample losses from these additional sample preparation steps can hinder sequence elucidation of less abundant peptidyl species.

Multistage MS of cyclized peptide can be used to avoid additional sample manipulation prior to MS. Multistage MS (primarily MS^{2-4}) can be harnessed to generate linear peptidyl species in the gas phase, where subsequent fragmentation proceeds via free N- and C-termini typical of linear peptides. Delocalized initial linearization requires sophisticated algorithms for the interpretation of MS^n experimental data.[131,133] Alternatively, gas-phase reactions that introduce sites (e.g. dehydroalanine) within the cyclic sequence that are favorable for fragmentation simplify analysis.[134] For example, recently developed gas-phase ion/ion reactions between cyclotides and sulfate radical anions within the ion optics of a mass spectrometer resulted in the conversion of select cyclotide cysteines to dehydroalanine.[134] Subsequent CID fragmentation generates site-specific linearized peptides, greatly reducing the downstream data complexity.[134] These gas-phases ion/ion reactions require non-standard, custom mass spectrometers, and may not be commercially available requiring extensive instrumentation experience to implement.

3.4 Oxidation (Met) / Hydroxylation (Trp, Tyr, Pro)

The addition of an oxygen or hydroxyl group commonly occur on AMPs with approximately the same frequency (Figure 2B). Hydroxylation of Trp, Tyr, and Pro is often biologically significant, while oxidation of Met is mainly attributed to sample handling (increased temperature, buffer conditions, ion source) and may result in decreased activity.[11,107,135,136] Multiple oxidative modifications can result in mixed peptidoform populations, producing chimeric fragmentation spectra that complicate sequencing efforts. Oxidative modifications can be identified using multistage MS strategies, including the detection of neutral losses and diagnostic ions.

Multistage MS. At the most basic level, oxidation and hydroxylation are recognized by identifying masses that are offset by +15.9949 Da,[137] whereby the number of modifications can be enumerated as multiples thereof. MS/MS analysis can provide secondary information to localize oxidative modifications. Peptides containing methionine sulfoxide readily generate a diagnostic and dominant neutral loss during fragmentation (Table 3), however this can preclude detection of additional product ions for sequencing.[62,138] In this case, the highly abundant neutral loss can be selected for MS^3 fragmentation to yield spectra with more informative peptide backbone fragmention.[138]

Certain hydroxylation modifications produce diagnostic ions. Immonium ions for hydroxytryptophan and hydroxytyrosine can be detected, but both have isobaric diamino acid ions (Val-Pro and Thr-Thr; Cys-Val, Asp-Ser, and Met-Ala, respectively) (Table 3) which interfere with confident identification.[61] Subsequent immonium ion fragmentation produces diagnostic masses for 5-hydroxytryptophan, 2-hydroxytryptophan, and 3-hydroxytyrosine (Table 3). These indicative peaks facilitate the identification of structural isomers and resolve isobaric masses. Another diagnostic ion (171.0674 *m/z*) was identified for hydroxyproline-containing peptides, corresponding to the hydroxyproline-glycine dipeptide *b*-ion.[139] This ion can be used to suggest the presence of hydroxyproline but is not diagnostic because the hydroxyproline-glycine motif is not universal to all hydroxyproline sites. It is also possible to discriminate between 3- and 4-hydroxyproline with *w*-ions containing an N-terminal hydroxyproline, where the 4-hydroxyproline containing *w*-ion retains the hydroxyl group and is detected +15.9949 Da from the equivalent 3-hydroxyproline-containing ion.[140,141] Resulting *w*-ions are most stable when there is a C-terminal basic residue, therefore this method is best applied to tryptic digests of AMPs.[140,141]

3.5 Glycosylation.

Currently, there are four known types of glycosylation, broadly classified by the sugar-peptide bond (*N-, O-, S-, C*-linked). Twelve AMPs contain at least one glycosylated residue, with *O*-linked glycosylation being the most common (Figure 2B).[8] The function of the glycan moiety ranges from stabilization to immunomodulation and can confer antimicrobial activity.[142] MS-based strategies to identify glycosylation include enzymatic deglycosylation and chemical cleavage. *N*-linked deglycosylation is most commonly pursued via Peptide-N-Glycosidase F (PNGase F) and/or Endoglycosidase H (Endo H) (Table 2). PNGase F is an amidase that cleaves *N*-linked glycans between Asn residues and the first sugar moiety, leaving both the core peptide and glycan intact.[52,143] Endo H also cleaves *N*-linked glycans, but leaves one N-acetylglucosamine residue on the peptide.[52,144] O-glycosidase removes *O*-linked core glycans from the peptide, however substituents on the glycan prevent its release and must be removed for effective *O*-glycosidase cleavage.[145,146] Hydrazinolysis can be used to remove both *N*- and *O*- linked glycans through

β-elimination (Table 2). Non-selective release can be achieved by incubation at high temperatures, while selective release of *O*-linked glycans requires milder conditions (60 ℃).[147] Hydrazinolysis leaves the glycan moiety intact but damages the peptide, resulting in minimal peptide sequence information.[148] After glycosylation has been identified it is non-trivial to characterize the specific glycan groups – the diverse glycoproteomics field has developed to predict, detect, and define glycosylations.[149–151]

3.6 Halogenation

Halogenation is a relatively rare modification (Figure 3B) that impacts peptide stability and activity.[8] The addition of bulky atoms tends to increase peptide stability by hindering peptidase accessibility.[152,153] Native NA-107, a lantipeptide produced by *Microbispora corallina*, contains a 5-chlorotryptophan which increases antimicrobial activity, possibly strengthening lipid binding interactions.[136,154] Rare, mono- and dibrominated Tyr and monobrominated Trp have been identified in AMPs from marine organisms.[152] Unlike common proteogenic elements (C, H, N, O, S), bromine and chlorine have two highly abundant isotopes (^{79}Br - 51 % and ^{81}Br - 49%; ^{35}Cl - 76 % and ^{37}Cl – 24 %) generating distinctive isotopic distributions which become more exaggerated with multiple halogenation modifications on the same peptide.[155,156] Database searching algorithms often incorrectly assign the M+2 peak as the monoisotopic mass, though the use of custom modifications can mitigate these effects.[157] Halogenated AMPs have been identified during manual interrogation of MS data based on their characteristic isotopic distribution.[158,159] Product ions containing the halogenated residue will retain the unusual distribution, facilitating PTM localization.[158,160,161] This strategy is useful for peptides under 5 kDa where the impact of halogenation on the isotopic distribution is most evident.[157,158]

3.7 Stereoisomers

Peptide stereochemistry can dramatically impact bioactivity,[63,162–164] but these same mass PTMs elude detection via mass spectrometry. For example, predominately included L-amino acids can be post-translationally isomerized to the D- form, producing a heterogeneous population of stereoisomers with different stabilities and/or activities.[63,163,164] *Cis/trans* isomerization of the peptide backbone via spontaneous or peptidyl-prolyl isomerase mediated mechanisms can also alter activity.[162,165] Although the *trans* conformation is thermodynamically favored, the sidechain cyclization of Pro reduces the energy barrier between stereoisomers resulting in a higher proportion of *cis* Xaa-Pro bonds, especially when X is another Pro or aromatic residue (Trp, Tyr, Phe).[162,166–169] Approaches to determine stereochemistry include Edman degradation (D-amino acids) and NMR (D-amino acids and *cis/trans* isomerization), but these require lengthy isolation steps to obtain sufficient purified material.[63] Here, we discuss enzymatic cleavage, separation, and multistage MS methods available to identify stereoisomers without prior isolation. Innovative MS-based approaches to differentiate stereoisomers comprise an exciting area of research with the potential to streamline AMP characterization.

Enzymatic cleavage: D-amino acid containing peptides (DAACPs) and peptides with *cis*-Pro are often resistant to proteases, most of which preferentially cleave *trans* L-amino acids.[163] Strategies have been developed to use this inherent enzymatic stability to identify DAACPs with a D-amino acid near the N-terminus. Peptides are incubated with aminopeptidase M and screened with MS to identify peptides recalcitrant to enzymatic digestion (Figure 5).[50,170,171] Aminopeptidase M will not cleave at N-terminal Asp, Glu, Pro, Xaa-Pro or modified N-termini.[50] As such, aminopeptidase digestions are most useful for

identifying D-amino acids which are near the N-terminus of AMPs whose primary sequence and post-translational modifications are already known.[50] Aminopeptidase M could be paired with other digestions to produce shorter peptides placing interior residues and C-terminal residues near the N-terminus of a digest peptide and more accessible for DAACP analysis.

Separations: Liquid chromatography, capillary electrophoresis, and gas-phase ion mobility can be used to differentiate stereoisomers based on retention time, though reversed-phase chromatography is the most common.[63,162,172,173] The retention time of native peptides can be compared to synthetic peptides with known modifications to clarify the specific modification present in the native AMP. Iterative comparisons to synthetic peptides often require costly synthesis of many peptide variants and tedious chromatographic method development (e.g. stationary phase, temperature, gradient). To limit the number of synthetic peptides needed, AMPs can be digested and the resulting shorter peptides compared to synthetic standards to localize the modification.

Ion mobility (IM) separations are less ubiquitous but offer short analysis times, high sensitivity, and orthogonal separation to reversed-phase chromatorgraphy.[174] Briefly, peptides ionized in the gas phase are separated based on collisional cross section and different conformations can be resolved. IM has been applied to *cis/trans*-Pro populations and D-amino acids in AMPs.[173,175] However, like chromatographic techniques, standards are needed to compare drift times and localize modifications. Post-fragmentation IM is an emerging experimental design which can be used to reduce the number of synthetic peptides required to detect and localize D-amino acids.[176] Native D-amino acid containing AMPs and a single L- synthetic analog are fragmented in the ion optics of a mass spectrometer and then separated with IM prior to detection. The D-amino acid modification can be localized because peptide fragments containing the D-amino acid will have different drift times than their all L-counterparts. Access to IM instrumentation is a factor in the implementation, though it has rapidly become commercially available on multiple MS vendor platforms.

Multistage MS: Multistage MS methods for stereoisomers rely on differences in the intensity of fragment ions, as opposed to the generation of unique fragments. DAACP and *cis* containing peptide stereoisomers demonstrate unique MS/MS fragmentation patterns from their all L- or *trans* counterparts.[177–182] DAACP analysis requires an analogous all L-synthetic peptide to identify the presence of D-amino acids.[178–182] MS/MS fragmentation spectra must be composed of single species because interfering ions can skew the comparison between isomers and lead to inaccurate conclusions. Chemical derivatization (e.g. metal-bound trimeric complex ions, acetylation) can be used to enhance chiral fragmentation patterns.[178,183]

3.8. Summary

AMPs contain wide variety of PTMs which increase sequence diversity and impact peptide activity and stability. Individual PTMs vary greatly in frequency within known AMPs (Figure 2B). Here, strategies to identify the most common PTMs and those with known impacts on bioactivity are highlighted. Certain PTMs (e.g. terminal modifications and cyclization) produce similar results with chemical derivatization and exopeptidase digestion, thus requiring iterative experiments to identify specific modifications (Figure 4). Stereochemical modifications are uniquely challenging because they do not alter peptide mass. Innovative strategies using separations and/or differential fragmentation patterns have been developed to resolve stereoisomers. Identifying PTMs in tandem with peptide sequencing greatly reduces possible sequences and increases the likelihood of accurate characterization.

4. Case studies: Gomesin and styelin D

Antimicrobial peptides are highly complex and the path to full molecular characterization may not be straightforward. Numerous factors, including sequence length, number/type of PTMs, and the presence of multiple peptidoforms increase the challenge. Here, two cases studies are presented to illustrate the strengths and challenges of the approaches described above for implementation to real AMPs. (1) First, we propose an alternative workflow to characterize the AMP gomesin, emphasizing how orthogonal experiments can be used to support MS-based sequencing while avoiding chromatographic method development necessary for isolation. (2) Then, we discuss styelin D, demonstrating that MS-based methods and genomic information can be paired to address AMPs with high molecular complexity.

Gomesin is a highly-modified 17 amino acid AMP composed of all L-amino acids isolated from the arachnid *Acanthoscurria gomesiana*, containing two disulfide bonds, a pyroglutamic acid, and C-terminal amidation (Figure 6A).[69] Gomesin was characterized using Edman degradation, pyroglutamate aminopeptidase digestion, reduction/alkylation of disulfide bonds, trypsin digestion, and activity comparisons with synthetic peptides varying in C-terminal amidation.[69] This required peptide isolation and multiple synthetic peptides to confirm PTMs based on differential bioactivity. Based on the approaches detailed herein, we propose an alternative MS-based sequencing workflow that would enable identification of all PTMs *without* synthesis of additional synthetic peptides or the need for gomesin isolation.

1. Cysteine reduction and alkylation – Mass shift indicates two disulfide bonds and linearization enhances MS/MS fragmentation.
2. Methylation – Absence of mass shift reveals that gomesin lacks acidic residues and a modified C-terminus.
3. MS/MS neutral loss – Abundant ammonium neutral loss suggests that the C – terminus is amidated.
4. Dimethylation – Mass shift indicates that gomesin contains a single primary amine, but not its associated sequence feature (lysine or N-terminus).
5. Aminopeptidase digestion – Gomesin remains intact after incubation with aminopeptidase M revealing that the N-terminus is modified and that single primary amine indicated by dimethylation must be a Lys. Further experiments are required to identify the specific N-terminal modification.
6. Pyroglutamate aminopeptidase - Gomesin is digested by pyroglutamate aminopeptidase and thus contains an N-terminal pyroglutamic acid.
7. *De novo* sequencing - Supplemental information about amino acid compositions (4 Cys, 0 Gly/Asp, 1 Lys) and PTM (two disulfide bonds, amidation, pyroglutamic acid) can be used during *de novo* sequencing to constrain possible sequences. After sequencing is completed, Leu/Ile and disulfide connectivity must still be resolved.
8. Multistage MS - Gomesin contains a single Leu/Ile, and is thus an ideal candidate for MS^3 differentiation.
9. Intact trypsin digestion - As in the originally published workflow, trypsin digestion of intact gomesin would reveal disulfide connectivity.[69]

Overall, gomesin is an example of a peptide that is well suited to sequence characterization with complementary MS-based methods because it contains PTMs which can be directly detected by chemical derivatization, enzymatic digestions, and multistage MS.

Styelin D is a 32 residue, all L- amino acid, peptide from *Styela clava* containing C-terminal amidation, 6-bromotryptophan, dihydroxyarginine, 3,4-dihydroxyphenylalanine, 5-hydroxylysine, and dihydroxylysine extracted as a mixture of peptidoforms varying in the extent and localization of lysine hydroxylation (Figure 6B).[11] Comparison between the activity of fully modified synthetic styelin D and native styelin D composed of several hydroxylation variants revealed that the native mixture of peptides was more active in acidic and high salt conditions.[11] Therefore, characterization of styelin D requires elucidation of all variants. Researchers used a combination of a cDNA library, Edman degradation, and mass spectrometry to elucidate the sequence variants of styelin D.[11] Ideally, mass spectrometry alone could be used to determine the primary sequence, modifications, and PTM localization; however, the heterogenous population and unusual PTMs make styelin D significantly more challenging to characterize than gomesin. Styelin D variants which contain the same number of hydroxylations but with different localization have the same exact mass, and are thus co-selected for fragmentation by the mass spectrometer. Co-fragmentation of multiple distinct peptides with the same mass results in chimeric MS/MS spectra that are far more complex and difficult to interpret than the spectra of each individual peptide (Figure 6C-E). Bromination results in a distinct isotopic distribution which facilitates the identification of brominated peptides but complicates sequencing efforts. Many *de novo* sequencing and database searching algorithms are ill suited for processing peptides with abnormal isotopic distributions and monoisotopic peaks may be mis-assigned during manual sequencing. The molecular complexity and unusual PTMs of styelin D create a situation where MS strategies alone are unlikely to produce the sequence solution.

Alternatively, MS-based techniques and *S. clava* genomic information could be leveraged together to develop an efficient sequencing strategy.

1. Generate *S. clava* protein database - Genomic or transcriptomic data can be translated *in silico* to generate a database of predicted proteins including styelin D.
2. Separation of styelin D variants – Same mass variants of steylin D could be chromatographically resolved with online reversed-phase liquid chromatography or ion mobility prior to MS analysis. Temporarily separated variants can be individual fragmented, avoiding chimeric MS/MS spectra.
3. Sequence tag - A series of sequential residues commonly identified within the fragmentation spectra of styelin D could be used to search for precursor peptides within the protein database. Correct proteolytic processing (e.g. removal of signal peptides) and PTMs (e.g. bromination and amidation) must still be identified for those moieties revealed.
4. Identify amino acids – Methods such as dimethyl labeling, methylation, alkylation, and enzymatic/chemical digestions would provide direct evidence of the number/type of unmodified amino acids in styelin D. This constrains the sequence of mature styelin D and limits feasible proteolytic processing of the precursor peptide.
5. Identify PTMs – Strategies to identify presence or absence of specific PTMs are equally important.For example, styelin D variants differ by increments of 16 Da, likely caused by oxidative modifications. Artificial Met oxidation is far more common than Try, Lys, or Arg hydroxylation, but the lack of an oxidized Met neutral loss in MS/MS spectra would prompt consideration of oxidative modifications that might otherwise be disregarded.

Styelin D is a challenging sequencing target whose molecular complexity (e.g. unusual modification and heterogenous population) may prohibit sequencing via MS only methods. Genomic/transcriptomic sequencing has become immensely cheap and accessible and can be leveraged for primary sequence

characterization where possible. However, post-translational transformations (i.e. proteolytic processing and PTMs) that confer activity but increase complexity are more readily addressed by MS-based approaches summarized herein. No single workflow will be effective for all peptides and each case must use a series of experiments to winnow down to the accurate native peptide with the most elegant and direct approaches to gain the maximum amount of data with the least amount of effort. Additional experiments (e.g. Edman sequencing, NMR, etc.), where necessary, can provide further constraints and/or orthogonal validation. Together, these two case studies illustrate how MS-based methods discussed herein have broad applicability to many sequencing workflows.

5. Peptidoform heterogeneity and functional implications

Peptidoforms, analogous to proteoforms, are peptide variants derived from a single gene,[184–186] often differentiated by PTM presence/localization. AMP peptidoforms have been observed in a wide variety of species including fish,[162] amphibians,[187] plants,[107] mammals,[12,13,188] insects,[15] and tunicates.[11] Despite this, the extent to which an organism expresses AMP peptidoforms is poorly understood and little is known about their specific structural/functional implications.

Peptidoform heterogeneity: LC-MS is well suited to rapidly profile AMP peptidoforms in crude extracts. Resultant data can be processed via manual interrogation and/or higher throughput bioinformatic strategies. A two-dimensional display of LC-MS data, where ions are plotted by retention time and *m/z* to reveal clusters of ions with mass differences that correspond to PTMs, can facilitate visualization of relationships between peptidoforms and facilitate manual identification.[187,189,190] Especially useful when an observed mass shift is readily associated with a known modification, related peptides may be missed if they differ by an unusual mass shift. Additionally, diagnostic MS/MS fingerprint ions can rapidly identify certain AMP classes prior to full sequence elucidation.[191,192]

Bioinformatic approaches to identify unknown or unexpected PTMs are rapidly developing and can be leveraged to reveal AMP peptidoforms. Most can be categorized as database searching or spectral networking methods. While major limitations exist for database searches to establish AMP primary sequence (as discussed previously), this can be a particularly useful approach to implement when identifying AMP peptidoforms with a known sequence. This can be paired with digestions, where necessary, to create shorter peptides more amenable to analysis. Two database searching paradigms that consider many possible modifications dominate the field: open (mass-tolerant)[193] and error-tolerant[194] searching. These searches can efficiently identify multiple peptidoforms because they consider all possible modifications without a bias towards user defined/expected PTMs. While powerful, these methods can result in false positives resulting from the increased search space compared to analyses with defined modifications.[195,196] Because manual validation of database matches is feasible for peptidoforms from a limited number AMPs, this informatics approach can be leveraged to rapidly identify putative unknown peptidoforms.

Spectral networking is complementary to database searching, does not require a protein database, and is well suited to the analysis of intact, fully processed AMPs.[20,29,37,197–199] This relies on the principle that very closely related peptides will have similar fragmentation spectra. Each spectrum in a dataset is a called a node and virtual edges connect related spectra produced by peptides that vary by a single amino acid mutation or modification to form spectral pairs. Multiple spectral pairs are connected to form spectral networks, and the other peptides within an AMP's spectral network represent potential peptidoforms. Spectral networking holds enormous potential to identify AMP peptidoforms as it relies on statistically

significant grouping of peptides rather than requiring the identification of each individual peptide during an error- or mass-tolerant database search.[197]

Functional implications: Discovery and characterization of AMP peptidoforms alone is not enough, and understanding the resultant impact on respective biological functions is critical. Although most AMPs have been shown to have minor structural variants, few have been more thoroughly studied to resolve differences in activity. For example, three insect-derived drosocin peptidoforms were found to vary in glycosylation and exhibit differential activity against Gram-negative bacteria.[15] Two peptidoforms of SAMP H1 were isolated from Atlantic salmon but only the variant containing a *cis*-proline was antibacterial.[162] Unsurprisingly, peptidoforms of human AMPs are the most extensively molecularly and functionally characterized. Human AMP peptidoforms impact target specificity, immune response and pathogen mechanism of resistance - emphasizing the importance of understanding the broader peptidoform structure / function landscape.

Redox modifications are common regulators of activity and AMPs contain many potential redox active sites, including Cys residues.[200] Disulfide bonding is often a hallmark characteristic of AMP families but can be dynamic; changes in disulfide bonding result in the creation of peptidoforms and can impact activity.[12,201–203] For example, variable oxidation states of the three disulfide bonds in human β-defensin 1 (hBD-1) modulate antimicrobial specificity and mechanism of action (MOA) based on environmental conditions.[12,204,205] Reduced hBD-1 is bactericidal and bacteriostatic to both Gram-positive and -negative bacteria, causing damage to bacterial membranes and entrapping bacteria in a net-like structure.[206] Oxidized hBD-1, with disulfide bonds intact, is active against only Gram-negative bacteria under aerobic conditions via an MOA that is not fully charactrized.[205] Observations that hBD-1 activity can be controlled by oxygen content and reductive potential of culture conditions has led to the hypothesis that HDB-1 activity can be regulated by its environment.

Peptidoform variants resulting in modification of charged residues can impact antimicrobial and immunomodulatory activity. Understanding how AMPs are modified and the impact of these modifications can reveal important host-pathogen interactions. Human neutrophil peptide-1 (HNP-1), a defensin, has been isolated with three peptidoforms: unmodified, mono-ADP-ribosylated, and di-ADP-ribosylated.[13,14] ADP-ribosylated HNP-1 has a lower net charge and antimicrobial/cytotoxic activity, but greater immunomodulatory activity.[207,208] It is hypothesized that host cells ribosylate HNP-1 as a regulatory mechanism so that it can be expressed at high concentrations that facilitate immunomodulation while circumventing harmful cytotoxic effects.[207]

AMPs interact directly with target species creating the possibility that pathogens can modify AMPs and modify peptidoforms as a mechanism of resistance. Interestingly, it appears possible in the case of ADP-ribosylation of HNP-1. *In vitro* assays show *E. coli* enzymes can ADP-ribosylate HNP-1 with the same efficiency as human enzymes resulting in a decrease in antibacterial activity.[209] Similarly, arginine residues in cathelicidin LL-37 can be citrullinated by rhinovirus to decrease net positive charge and diminish antiviral activity.[188] These studies suggest that specific AMP peptidoforms may be advantageous to either the host or pathogen, emphasizing the importance of understanding the biological impacts of different AMP peptidoforms and mechanisms of resistance evolution.

Although new AMPs are discovered each year, definition of native peptidoform heterogeneity remains underexplored. This can be much improved through more thorough examination of the resultant LC-MS data as described. It will be imperative moving forward to apply innovative approaches to discern the roles/relevance of specific peptidoforms in the biological context.

6. Conclusions

Antimicrobial peptides are an exciting class of bioactive natural products with tremendous chemical diversity further complicated by high heterogeneity. Rapid and thorough AMP characterization, including peptidoform profiling, is essential to discovering new molecular species and understanding their potential biological roles.

Mass spectrometric approaches can streamline AMP characterization, though often through non-standard experimental workflows. Peptide size and diverse PTMs generate enormous theoretical sequence space. While computational approaches continue to evolve to meet the needs of the AMP community,[29,32–36] orthogonal experimental approaches to reveal amino acid and PTMs composition, such as those reviewed herein, can be used as constraints in sequencing algorithms and/or provide critical clues when manually sequencing.

AMPs with common modifications, such as amidation or disulfide bonding, are well studied for their broader biological relevance and has resulted in established methods for their characterization. Even so, it is essential to appreciate the vast diversity of AMPs and consider a wider range of possible modifications. Peptides containing unusual modifications, such as halogenation or D-amino acids, may have increased proteolytic stability or other characteristics that make them extremely attractive lead compounds.[50,153] AMPs with extreme structural complexity highlight that peptide molecular characterization requires a broad toolbox of strategies which includes mass spectrometry, DNA/RNA sequencing, etc. Optimal characterization workflows must be developed for individual peptides – which can be more efficiently refined with knowledge of amino acid and PTM composition.

Advances in proteomics continue to reveal new PTMs emphasizing that the current understanding of proteomic and peptidomic diversity is incomplete.[210] As such, it can be expected that AMP sequence diversity will continue to expand as new peptides are discovered. Likewise, new and innovative methods that address newly discovered peptide chemical space must be integrated into AMP workflows.

Conflicts of Interest

There are no conflicts of interest to declare.

Acknowledgements

This work was funded by NIH (1R01GM125814) to L.M.H. T.B.M. was supported by the NSF Graduate Research Fellowship program (DGE-1650016) and N.C.P. by the NIH Biophysics training grant (T32 GM008570) and the UNC Graduate School Dissertation Completion Fellowship. The authors thank Guangshun Wang (University of Nebraska Medical Center) for supplying access to the Antimicrobial Peptide Database entries.

References

1 P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K.-D. Entian, M. A. Fischbach, J. S. Garavelli, U. Göransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Müller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L.

Patchett, J. Piel, M. J. T. Reaney, S. Rebuffat, R. P. Ross, H.-G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Süssmuth, J. R. Tagg, G.-L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108–60.

2 M. Zasloff, *Nature*, 2002, **415**, 389–395.

3 H. Jenssen, P. Hamill and R. E. W. Hancock, *Clin. Microbiol. Rev.*, 2006, **19**, 491–511.

4 F. Guilhelmelli, N. Vilela, P. Albuquerque, L. da S. Derengowski, I. Silva-Pereira and C. M. Kyaw, *Front. Microbiol.*, 2013, **4**, 353.

5 L. D. S. Machado, E. V. B. de Carvalho, F. V. de A. E. Silva, P. V. D. S. Cabreira and O. L. Franco, *Curr. Top. Med. Chem.*, 2017, **17**, 520–536.

6 E. F. Haney, S. K. Straus and R. E. W. Hancock, *Front. Chem.*, 2019, **7**, 43.

7 D. Gaspar, A. S. Veiga and M. A. R. B. Castanho, *Front. Microbiol.*, 2013, **4**, 294.

8 G. Wang, X. Li and Z. Wang, *Nucleic Acids Res.*, 2016, **44**, D1087–D1093.

9 L. T. Nguyen, E. F. Haney and H. J. Vogel, *Trends Biotechnol.*, 2011, **29**, 464–472.

10 M. Mahlapuu, J. Håkansson, L. Ringstad and C. Björn, *Front. Cell. Infect. Microbiol.*, 2016, **6**, 194.

11 S. W. Taylor, A. G. Craig, W. H. Fischer, M. Park and R. I. Lehrer, *J. Biol. Chem.*, 2000, **275**, 38417–38426.

12 B. O. Schroeder, Z. Wu, S. Nuding, S. Groscurth, M. Marcinowski, J. Beisner, J. Buchner, M. Schaller, E. F. Stange and J. Wehkamp, *Nature*, 2011, **469**, 419–423.

13 G. Paone, L. A. Stevens, R. L. Levine, C. Bourgeois, W. K. Steagall, B. R. Gochuico and J. Moss, *J. Biol. Chem.*, 2006, **281**, 17054–17060.

14 E. Balducci, A. Bonucci, M. Picchianti, R. Pogni and E. Talluri, *Int. J. Pept.*, 2011, **2011**, 1–7.

15 D. S. Lele, G. Kaur, M. Thiruvikraman and K. J. Kaur, *Glycoconj. J.*, 2017, **34**, 613–624.

16 H. Mohimani, R. D. Kersten, W.-T. Liu, M. Wang, S. O. Purvine, S. Wu, H. M. Brewer, L. Pasa-Tolic, N. Bandeira, B. S. Moore, P. A. Pevzner and P. C. Dorrestein, *ACS Chem. Biol.*, 2014, **9**, 1545–1551.

17 L.-F. Nothias, M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A. Sarvepalli, P. Leyssen, D. Touboul, J. Costa, J. Paolini, T. Alexandrov, M. Litaudon and P. C. Dorrestein, *J. Nat. Prod.*, 2018, **81**, 758–767.

18 M. A. Skinnider, C. W. Johnston, R. E. Edgar, C. A. Dejong, N. J. Merwin, P. N. Rees and N. A. Magarvey, *Proc. Natl. Acad. Sci.*, 2016, **113**, E6343–E6351.

19 V. Tracanna, A. de Jong, M. H. Medema and O. P. Kuipers, *FEMS Microbiol. Rev.*, 2017, **41**, 417–429.

20 L. Cao, A. Gurevich, K. L. Alexander, C. B. Naman, T. Leão, E. Glukhov, T. Luzzatto-Knaan, F. Vargas, R. Quinn, A. Bouslimani, L. F. Nothias, N. K. Singh, J. G. Sanders, R. A. S. Benitez, L. R. Thompson, M. N. Hamid, J. T. Morton, A. Mikheenko, A. Shlemov, A. Korobeynikov, I. Friedberg, R. Knight, K. Venkateswaran, W. H. Gerwick, L. Gerwick, P. C. Dorrestein, P. A. Pevzner and H. Mohimani, *Cell Syst.*, 2019, **9**, 600-608.e4.

21 R. Hammami, J. Ben Hamida, G. Vergoten and I. Fliss, *Nucleic Acids Res.*, 2009, **37**, D963–D968.

22 G. Wang, X. Li and Z. Wang, *Nucleic Acids Res.*, 2009, **37**, D933-937.

23 Z. Wang and G. Wang, *Nucleic Acids Res.*, 2004, **32**, D590–D592.

24 J. P. Mulvenna, C. Wang and D. J. Craik, *Nucleic Acids Res.*, 2006, **34**, D192–D194.

25 C. K. Wang, Q. Kaas, L. Chiche and D. J. Craik, *Nucleic Acids Res.*, 2008, **36**, D206–D210.

26 M. Mann, *Clin. Chem.*, 2016, **62**, 293–4.

27 P. Edman, *Arch. Biochem.*, 1949, **22**, 475–476.

28 R. Deutzmann, in *Molecular Diagnosis of Infectious Diseases*, 2004, pp. 269–297.

29 H. Mohimani and P. A. Pevzner, *Nat. Prod. Rep.*, 2016, **33**, 73–86.

30 M. Mann and N. L. Kelleher, *Proc. Natl. Acad. Sci.*, 2008, **105**, 18132–18138.

31 L. A. Macias, I. C. Santos and J. S. Brodbelt, *Anal. Chem.*, 2020, **92**, 227–251.

32 G. Menschaert and D. Fenyö, *Mass Spectrom. Rev.*, 2017, **36**, 584–599.

33 J. Armengaud, J. Trapp, O. Pible, O. Geffard, A. Chaumot and E. M. Hartmann, *J. Proteomics*, 2014, **105**, 5–18.

34 K. F. Medzihradszky and R. J. Chalkley, *Mass Spectrom. Rev.*, 2015, **34**, 43–63.

35 T. Muth, F. Hartkopf, M. Vaudel and B. Y. Renard, *Proteomics*, 2018, **18**, 1700150.

36 J. Allmer, *Expert Rev. Proteomics*, 2011, **8**, 645–657.

37 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D.-T. T. D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. C. Liaw, Y.-L. L. Yang, H.-U. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. M. Allard, P. Phapale, L.-F. F. Nothias, T. Alexandrov, M. Litaudon, J.-L. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. T. D. D. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, C. A. P. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane,

J. Gurr, A. M. C. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. M. Allard, P. Phapale, L.-F. F. Nothias, T. Alexandrov, M. Litaudon, J.-L. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. T. D. D. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking*, Nature Publishing Group, 2016, vol. 34.

38 T. Müller and D. Winter, *Mol. Cell. Proteomics*, 2017, **16**, 1173–1187.

39 S. Suttapitugsakul, H. Xiao, J. Smeekens and R. Wu, *Mol. Biosyst.*, 2017, **13**, 2574–2582.

40 J. R. Kramer and T. J. Deming, *Biomacromolecules*, 2012, **13**, 1719–1723.

41 J.-L. Hsu, S.-Y. Huang, J.-T. Shiea, W.-Y. Huang and S.-H. Chen, *J. Proteome Res.*, 2005, **4**, 101–108.

42 D. R. Goodlett, A. Keller, J. D. Watts, R. Newitt, E. C. Yi, S. Purvine, J. K. Eng, P. von Haller, R. Aebersold and E. Kolker, *Rapid Commun. Mass Spectrom.*, 2001, **15**, 1214–1221.

43 N. I. Martin, T. Sprules, M. R. Carpenter, P. D. Cotter, C. Hill, R. P. Ross and J. C. Vederas, *Biochemistry*, 2004, **43**, 3049–3056.

44 R. Joppich-Kuhn, J. A. Corkill and R. W. Giese, *Anal. Biochem.*, 1982, **119**, 73–77.

45 L. Tsiatsiani and A. J. R. Heck, *FEBS J.*, 2015, **282**, 2612–2626.

46 G. R. Dhapeau, Y. Boilp and J. Houmard, *J. Biol. Chem.*, 1972, **247**, 6720–6726.

47 S. J. Bark, N. Muster, J. R. Yates and G. Siuzdak, *J. Am. Chem. Soc.*, 2001, **123**, 1774–1775.

48 K. Tanabe, A. Taniguchi, T. Matsumoto, K. Oisaki, Y. Sohma and M. Kanai, *Chem. Sci.*, 2014, **5**, 2747.

49 R. P. Ambler, *Methods Enzymol.*, 1972, **25**, 143–154.

50 I. Livnat, H.-C. Tai, E. T. Jansson, L. Bai, E. V. Romanova, T. Chen, K. Yu, S. Chen, Y. Zhang, Z. Wang, D. Liu, K. R. Weiss, J. Jing and J. V. Sweedler, *Anal. Chem.*, 2016, **88**, 11868–11876.

51 E. Fowler, M. Moyer, R. G. Krishna, C. C. Q. Chin and F. Wold, in *Current Protocols in Protein Science*, 2001, vol. 3, p. Chapter 11.

52 L. Cao, J. K. Diedrich, Y. Ma, N. Wang, M. Pauthner, S. K. R. Park, C. M. Delahunty, J. S. McLellan, D. R. Burton, J. R. Yates and J. C. Paulson, *Nat. Protoc.*, 2018, **13**, 1196–1212.

53 T. P. Patel and R. B. Parekh, *Methods Enzymol.*, 1994, **230**, 57–66.

54 A. M. Falick, W. M. Hines, K. F. Medzihradszky, M. A. Baldwin and B. W. Gibson, *J. Am. Soc. Mass Spectrom.*, 1993, **4**, 882–893.

55 M. Fälth, M. M. Savitski, M. L. Nielsen, F. Kjeldsen, P. E. Andren and R. A. Zubarev, *Anal. Chem.*, 2008, **80**, 8089–8094.

56 H. J. Cooper, R. R. Hudgins, K. Håkansson and A. G. Marshall, *J. Am. Soc. Mass Spectrom.*, 2002, **13**, 241–249.

57 Q. Xia, M. V. Lee, C. M. Rose, A. J. Marsh, S. L. Hubler, C. D. Wenger and J. J. Coon, *J. Am. Soc. Mass Spectrom.*, 2011, **22**, 255–264.

58 A. Armirotti, E. Millo and G. Damonte, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 57–63.

59 R. Krüger, C.-W. Hung, M. Edelson-Averbukh and W. D. Lehmann, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 1709–1716.

60 L. Mouls, G. Subra, J.-L. Aubagnac, J. Martinez and C. Enjalbal, *J. Mass Spectrom.*, 2006, **41**, 1470–1483.

61 L. Mouls, E. Silajdzic, N. Haroune, C. M. Spickett and A. R. Pitt, *Proteomics*, 2009, **9**, 1617–1631.

62 F. M. Lagerwerf, M. van de Weert, W. Heerma and J. Haverkamp, *Rapid Commun. Mass Spectrom.*, 1996, **10**, 1905–1910.

63 E. T. Jansson, *J. Sep. Sci.*, 2018, **41**, 385–397.

64 Uniprot Consortium, *Nucleic Acid Res.*, 2019, **47**, D506–D515.

65 L. J. Hohmann, J. K. Eng, A. Gemmill, J. Klimek, O. Vitek, G. E. Reid and D. B. Martin, *Anal. Chem.*, 2008, **80**, 5596–606.

66 J. A. Madsen, D. R. Boutz and J. S. Brodbelt, *J. Proteome Res.*, 2010, **9**, 4205–4214.

67 J. N. deGruyter, L. R. Malins and P. S. Baran, *Biochemistry*, 2017, **56**, 3863–3873.

68 A. Miseta and P. Csutora, *Mol. Biol. Evol.*, 2000, **17**, 1232–1239.

69 P. I. Silva, S. Daffre and P. Bulet, *J. Biol. Chem.*, 2000, **275**, 33464–33470.

70 D. C. Ireland, R. J. Clark, N. L. Daly and D. J. Craik, *J. Nat. Prod.*, 2010, **73**, 1610–1622.

71 C. Giglione, A. Boularot and T. Meinnel, *Cell. Mol. Life Sci.*, 2004, **61**, 1455–1474.

72 G. K. T. Nguyen, W. H. Lim, P. Q. T. Nguyen and J. P. Tam, *J. Biol. Chem.*, 2012, **287**, 17598–17607.

73 D. S. Radchenko, S. Kattge, S. Kara, A. S. Ulrich and S. Afonin, *Biochim. Biophys. Acta - Biomembr.*, 2016, **1858**, 2019–2027.

74 J. M. Froelich and G. E. Reid, *Proteomics*, 2008, **8**, 1334–1345.

75 V. N. Lapko, D. L. Smith and J. B. Smith, *J. Mass Spectrom.*, 2000, **35**, 572–575.

76 H. G. Gundlach, S. Moore and W. H. Stein, *J. Biol. Chem.*, 1959, **234**, 1761–1764.

77 J. Zang, Y. Chen, W. Zhu and S. Lin, *Biochemistry*, 2020, **59**, 132–138.

78 C. W. Yeung, F. H. Carpenter and W.-D. Busse, *Biochemistry*, 1977, **16**, 1635–1641.

79 G. Diamond, M. Zasloff, H. Eck, M. Brasseur, W. Lee Maloy and C. L. Bevins, *Proc. Nati. Acad. Sci. USA*, 1991, **88**, 3952–3956.

80 M. Beukes, G. Bierbaum, H. Sahl and J. W. Hastings, *Appl. Environ. Microbiol.*, 2000, **66**, 23–28.

81 A. Axen, A. Carlsson I, A. Engstrom and H. Bennich, *Eur. J. Biochem*, 1997, **247**, 614–619.

82 R. Kaiser and L. Metzka, *Anal. Biochem.*, 1999, **266**, 1–8.

83 K. Srzentić, K. O. Zhurov, A. A. Lobas, G. Nikitin, L. Fornelli, M. V. Gorshkov and Y. O. Tsybin, *J. Proteome Res.*, 2018, **17**, 2005–2016.

84 New Jersey Department of Health and Senior Services, *Hazardous Substance Fact Sheet: Cyanogen bromide*, 2004.

85 S. R. Dennison, F. Harris, M. Mura and D. A. Phoenix, *Curr. Protein Pept. Sci.*, 2018, **19**, 823–838.

86 F. Harris, S. Dennison and D. Phoenix, *Curr. Protein Pept. Sci.*, 2009, **10**, 585–606.

87 S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas and H. B. Bode, *Org. Biomol. Chem.*, 2011, **9**, 3130–3132.

88 J. K. Seo, J. Stephenson, J. M. Crawford, K. L. Stone and E. J. Noga, *Mar. Biotechnol.*, 2010, **12**, 543–551.

89 P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed and A. J. R. Heck, *Nat. Protoc.*, 2009, **4**, 484–494.

90 M. Castro, R. Matsushita, A. Sebben, M. Sousa and W. Fontes, *Protein Pept. Lett.*, 2005, **12**, 89–93.

91 Mingming Ma, A. Kimberly K. Kutz-Naber and L. Li, *Anal. Chem.*, 2007, **79**, 673–681.

92 A. K. Mandal, M. Ramakrishnan, S. Ramasamy, V. Sabareesh, M. E. Openshaw, K. S. Krishnan and P. Balaram, *J Am Soc Mass Spectrom*, 2007, **18**, 1396–1404.

93 R. Liao, Y. Gao, M. Chen, L. Li and X. Hu, *Anal. Chem.*, 2018, **90**, 13533–13540.

94 O. Kleifeld, A. Doucet, A. Prudova, U. Auf Dem Keller, M. Gioia, J. N. Kizhakkedathu and C. M. Overall, *Nat. Protoc.*, 2011, **6**, 1578–1611.

95 J. L. Hsu and S. H. Chen, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, , DOI:10.1098/rsta.2015.0364.

96 M. Miyashita, A. Kitanaka, M. Yakio, Y. Yamazaki, Y. Nakagawa and H. Miyagawa, *Toxicon*, 2017, **139**, 1–12.

97 R. H. Tailor, D. P. Acland, S. Attenborough, B. P. Cammue, I. J. Evans, R. W. Osborn, J. A. Ray, S. B. Rees and W. F. Broekaert, *J. Biol. Chem.*, 1997, **272**, 24480–7.

98 G. Fimland, K. Sletten and J. Nissen-Meyer, *Biochem. Biophys. Res. Commun.*, 2002, **295**, 826–827.

99 M. Sitbon, L. D'Auriol, H. Ellerbrok, C. André, J. Nishio, S. Perryman, F. Pozo, S. F. Hayes, K. Wehrly and P. Tambourin, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 5932–5936.

100 C. N. Poston, R. E. Higgs, J. You, V. Gelfanova, J. E. Hale, M. D. Knierman, R. Siegel and J. A. Gutierrez, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 1228–1236.

101 R. Kerfah, O. Hamelin, J. Boisbouvier and D. Marion, *J. Biomol. NMR*, 2015, **63**, 389–402.

102 K. H. Gardner and L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 357–406.

103 S. S. Zhokhov, S. V. Kovalyov, T. Y. Samgina and A. T. Lebedev, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1600–1611.

104 Y. Xiao, M. M. Vecchi and D. Wen, *Anal. Chem.*, 2016, **88**, 10757–10766.

105 K. Ambihapathy, T. Yalcin, H.-W. Leung and A. G. Harrison, *J. Mass Spectrom.*, 1997, **32**, 209–

215.

106 A. T. Lebedev, E. Damoc, A. A. Makarov and T. Y. Samgina, *Anal. Chem.*, 2014, **86**, 7017–7022.

107 T. B. Moyer, L. R. Heil, C. L. Kirkpatrick, D. Goldfarb, W. A. Lefever, N. C. Parsley, A. J. Wommack and L. M. Hicks, *J. Nat. Prod.*, 2019, **82**, 2744–2753.

108 T. Y. Samgina, S. V. Kovalev, M. D. Tolpina, P. Trebse, G. Torkar and A. T. Lebedev, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 842–852.

109 T. Nakamura, H. Nagaki, Y. Ohki and T. Kinoshita, *Anal. Chem.*, 1990, **62**, 311–313.

110 B. V. Popp and Z. T. Ball, *Chem. Sci.*, 2011, **2**, 690–695.

111 J. Ni and M. Kanai, *Top Curr Chem*, 2016, **372**, 103–124.

112 J. T. J. Cheng, J. D. Hale, J. Kindrachuk, H. Jessen, M. Elliott, R. E. W. Hancock and S. K. Straus, *Biophys. J.*, 2010, **99**, 2926–2935.

113 D. Wellner, C. Panneerselvam and B. L. Horecker, *Proc. Natl. Acad. Sci.*, 1990, **87**, 1947–1949.

114 D. Chelius, Kay Jing, Alexis Lueras, Douglas S. Rehder, Thomas M. Dillon, Alona Vizel, Rahul S. Rajan, Tiansheng Li, A. Michael J. Treuheit and P. V. Bondarenko, *Anal. Chem.*, 2006, **78**, 2370–2376.

115 H. Steiner, D. Hultmark, Å. Engström, H. Bennich and H. G. Boman, *Nature*, 1981, **292**, 246–248.

116 I. H. Lee, C. Zhao, Y. Cho, S. S. . Harwig, E. L. Cooper and R. I. Lehrer, *FEBS Lett.*, 1997, **400**, 158–162.

117 G. A. Birkemo, T. Lüders, Ø. Andersen, I. F. Nes and J. Nissen-Meyer, *Biochim. Biophys. Acta*, 2003, **1646**, 207–215.

118 A. C. Nascimento, L. C. Zanotta, C. M. Kyaw, E. N. F. Schwartz, C. A. Schwartz, A. Sebben, M. V. Sousa, W. Fontes and M. S. Castro, *Protein J.*, 2004, **23**, 501–508.

119 X. Xi, R. Li, Y. Jiang, Y. Lin, Y. Wu, M. Zhou, J. Xu, L. Wang, T. Chen and C. Shaw, *Biochimie*, 2013, **95**, 1288–1296.

120 K. Konno, M. Hisada, H. Naoki, Y. Itagaki, R. Fontana, M. Rangel, J. S. Oliveira, M. P. dos S. Cabrera, J. R. Neto, I. Hide, Y. Nakata, T. Yasuhara and T. Nakajima, *Peptides*, 2006, **27**, 2624–2631.

121 S. Tsunasawa and K. Narita, *Methods Enzymol.*, 1976, **45**, 552–61.

122 A. A. Vassilevski, S. A. Kozlov, T. A. Egorov and E. V. Grishin, *Methods Mol. Biol.*, 2010, **615**, 87–100.

123 B. P. Cammue, M. F. De Bolle, F. R. Terras, P. Proost, J. Van Damme, S. B. Rees, J. Vanderleyden and W. F. Broekaert, *J. Biol. Chem.*, 1992, **267**, 2228–2233.

124 J. Li, H. Yu, X. Xu, X. Wang, D. Liu and R. Lai, *Genomics*, 2007, **89**, 413–418.

125 J. C. Slootweg, E. F. Van Herwerden, M. F. M. Van Doremalen, E. Breukink, R. M. J. Liskamp and D. T. S. Rijkers, *Org. Biomol. Chem.*, 2015, **13**, 5997–6009.

126 D. E. Palmer, C. Pattaroni, K. Nunami, R. K. Chadha, M. Goodman, T. Wakamiya, K. Fukase, S. Horimoto, M. Kitazawa, H. Fujita, A. Kubo and T. Shiba, *J. Am. Chem. Soc.*, 1992, **114**, 5634–

5642.

127 C. T. Lohans and J. C. Vederas, *J. Antibiot. (Tokyo).*, 2014, **67**, 23–30.

128 D. J. Craik, N. L. Daly, T. Bond and C. Waine, *J. Mol. Biol.*, 1999, **294**, 1327–1336.

129 P. J. Knerr and W. A. van der Donk, *Annu. Rev. Biochem.*, 2012, **81**, 479–505.

130 J. D. Hegemann, M. Zimmermann, X. Xie and M. A. Marahiel, *Acc. Chem. Res.*, 2015, **48**, 1909–1919.

131 N. C. Parsley, C. L. Kirkpatrick, C. M. Crittenden, J. G. Rad, D. W. Hoskin, J. S. Brodbelt and L. M. Hicks, *Phytochemistry*, 2018, **152**, 61–70.

132 S. K. Sze, W. Wang, W. Meng, R. Yuan, T. Guo, Y. Zhu and J. P. Tam, *Anal. Chem.*, 2009, **81**, 1079–1088.

133 H. Mohimani, Y. L. Yang, W. T. Liu, P. W. Hsieh, P. C. Dorrestein and P. A. Pevzner, *Proteomics*, 2011, **11**, 3642–3650.

134 D. J. Foreman, N. C. Parsley, J. T. Lawler, U. K. Aryal, L. M. Hicks and S. A. McLuckey, *Anal. Chem.*, 2019, **91**, 15608–15616.

135 I. K. Ø. Hansen, J. Isaksson, A. G. Poth, K. Hansen, A. J. C. Andersen, C. S. M. Richard, H. M. Blencke, K. Stensvåg, D. J. Craik and T. Haug, *Mar. Drugs*, 2020, **18**, 1–20.

136 S. I. Maffioli, M. Iorio, M. Sosio, P. Monciardini, E. Gaspari and S. Donadio, *J. Nat. Prod.*, 2014, **77**, 79–84.

137 K. L. Schey and E. L. Finley, *Acc Chem Res*, 2000, **33**, 299–306.

138 Z. Guan, N. A. Yates and R. Bakhtiar, *J. Am. Soc. Mass Spectrom.*, 2003, **14**, 605–613.

139 D. P. Zolg, M. Wilhelm, T. Schmidt, G. Médard, J. Zerweck, T. Knaute, H. Wenschuh, U. Reimer, K. Schnatbaum and B. Kuster, *Mol. Cell. Proteomics*, 2018, **17**, 1850–1863.

140 D. B. Kassel and K. Biemann, *Anal. Chem.*, 1990, **62**, 1691–1695.

141 F. Ma, R. Sun, D. M. Tremmel, S. D. Sackett, J. Odorico and L. Li, *Anal. Chem.*, 2018, **90**, 5857–5864.

142 N. G. Bednarska, B. W. Wren and S. J. Willcocks, *Drug Discov. Today*, 2017, **22**, 919–926.

143 M.-S. Kim and D. Leahy, *Methods Enzymol.*, 2013, **533**, 259–263.

144 F. Maley, R. B. Trimble, A. L. Tarentino and T. H. Plummer, *Anal. Biochem.*, 1989, **180**, 195–204.

145 V. P. Bhavanandan, J. Umemoto and E. A. Davidson, *Biochem. Biophys. Res. Commun.*, 1976, **70**, 738–745.

146 H. Iwase and K. Hotta, *Methods Mol. Biol.*, 1993, 14, 151–159.

147 T. Patel, J. Bruce, A. Merry, C. Bigge, M. Wormald, R. Parekh and A. Jaques, *Biochemistry*, 1993, **32**, 679–693.

148 S. Takasaki, T. Mizuochi and A. Kobata, *Methods Enzymol.*, 1982, **83**, 263–268.

149 R. D. Kersten, N. Ziemert, D. J. Gonzalez, B. M. Duggan, V. Nizet, P. C. Dorrestein and B. S.

Moore, *Proc. Natl. Acad. Sci.*, 2013, **110**, E4407–E4416.

150 Y. Yang, V. Franc and A. J. R. Heck, *Trends Biotechnol.*, 2017, **35**, 598–609.

151 K. K. Palaniappan and C. R. Bertozzi, *Chem. Rev.*, 2016, **116**, 14277–14306.

152 A. Falanga, L. Lombardi, G. Franci, M. Vitiello, M. R. Iovene, G. Morelli, M. Galdiero and S. Galdiero, *Int. J. Mol. Sci.*, 2016, **17**, 1–18.

153 M. A. Funk and W. A. Van Der Donk, *Acc. Chem. Res.*, 2017, **50**, 1577–1586.

154 F. Castiglione, A. Lazzarini, L. Carrano, E. Corti, I. Ciciliato, L. Gastaldo, P. Candiani, D. Losi, F. Marinelli, E. Selva and F. Parenti, *Chem. Biol.*, 2008, **15**, 22–31.

155 K. K. Palaniappan, A. A. Pitcher, B. P. Smart, D. R. Spiciarich, A. T. Iavarone and C. R. Bertozzi, *ACS Chem. Biol.*, 2011, **6**, 829–836.

156 R. Oh, M. J. Lee, Y.-O. Kim, B.-H. Nam, H. J. Kong, J.-W. Kim, J. Y. Park, J.-K. Seo and D.-G. Kim, *Fish Shellfish Immunol.*, 2018, **83**, 425–435.

157 H. Liu, C. F. Lichti, B. Mirfattah, J. Frahm and C. L. Nilsson, *J. Proteome Res.*, 2013, **12**, 4248–4254.

158 S. S. Nair, C. L. Nilsson, M. R. Emmett, T. M. Schaub, K. H. Gowd, S. S. Thakur, K. S. Krishnan, P. Balaram and A. G. Marshall, *Anal. Chem.*, 2006, **78**, 8082–8088.

159 J. A. Jakubowski, W. P. Kelley and J. V. Sweedler, *Toxicon*, 2006, **47**, 688–699.

160 M. Vijayasarathy and P. Balaram, *Toxicon*, 2018, **144**, 68–74.

161 C. Li, T. Haug, M. K. Moe, O. B. Styrvold and K. Stensvåg, *Dev. Comp. Immunol.*, 2010, **34**, 959–968.

162 T. Lüders, G. Alice Birkemo, J. Nissen-Meyer, Ø. Andersen and I. F. Nes, *Antimicrob. Agents Chemother.*, 2005, **49**, 2399–2406.

163 G. Genchi, *Amino Acids*, 2017, **49**, 1521–1533.

164 P. D. Cotter, P. M. O'Connor, L. A. Draper, E. M. Lawton, L. H. Deegan, C. Hill and R. P. Ross, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18584–18589.

165 B. M. Dunyak and J. E. Gestwicki, *J. Med. Chem.*, 2016, **59**, 9622–9644.

166 W. J. Wedemeyer, E. Welker and H. A. Scheraga, *Biochemistry*, 2002, **41**, 14637–14644.

167 W.-J. Wu and D. P. Raleigh, *Biopolymers*, 1998, **45**, 381–394.

168 B. Dasgupta, P. Chakrabarti and G. Basu, *FEBS Lett.*, 2007, **581**, 4529–4532.

169 H. K. Ganguly, B. Majumder, S. Chattopadhyay, P. Chakrabarti and G. Basu, *J. Am. Chem. Soc.*, 2012, **134**, 4661–4669.

170 P. C. Montecucchi, R. de Castiglione, S. Piani, L. Gozzini and V. Erspamer, *Int. J. Pept. Protein Res.*, 1981, **17**, 275–83.

171 H. C. Tai, J. W. Checco and J. V. Sweedler, in *Methods in Molecular Biology*, Humana Press Inc., 2018, vol. 1719, pp. 107–118.

172 W. A. Tao, L. Wu and R. G. Cooks, *Chem. Commun.*, 2000, 2023–2024.

173 K. Jeanne Dit Fouque, J. D. Hegemann, S. Zirah, S. Rebuffat, E. Lescop and F. Fernandez-Lima, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 1038–1045.

174 M. A. Ewing, M. S. Glover and D. E. Clemmer, *J. Chromatogr. A*, 2016, **1439**, 3–25.

175 M. T. Q. de Magalhães, E. A. Barbosa, M. V Prates, R. M. Verly, V. H. O. Munhoz, I. E. de Araújo, C. Bloch and Jr, *PLoS One*, 2013, **8**, e59255.

176 C. Jia, C. B. Lietz, Q. Yu and L. Li, *Anal. Chem.*, 2014, **86**, 2972–2981.

177 S. Warnke, C. Baldauf, M. T. Bowers, K. Pagel and G. Von Helden, *J. Am. Chem. Soc.*, 2014, **136**, 10308–10314.

178 E. Sachon, G. Clodic, C. Galanth, M. Amiche, C. Ollivaux, D. Soyez and G. Bolbach, *Anal. Chem.*, 2009, **81**, 4389–4396.

179 J. Koehbach, C. W. Gruber, C. Becker, D. P. Kreil and A. Jilek, *J. Proteome Res.*, 2016, **15**, 1487–96.

180 P. P. Hurtado and P. B. O'Connor, *Mass Spectrom. Rev.*, 2012, **31**, 609–625.

181 S. V. Serafin, R. Maranan, K. Zhang and T. H. Morton, *Anal. Chem.*, 2005, **77**, 5480–5487.

182 H. Awad and A. El-Aneed, *Mass Spectrom. Rev.*, 2013, **32**, 466–483.

183 W. A. Tao and R. G. Cooks, *Angew. Chemie Int. Ed.*, 2001, **40**, 757–760.

184 T. K. Toby, L. Fornelli and N. L. Kelleher, *Annu. Rev. Anal. Chem.*, 2016, **9**, 499–519.

185 L. M. Smith and N. L. Kelleher, *Nat. Methods*, 2013, **10**, 186–187.

186 G. Rosenberger, Y. Liu, H. L. Röst, C. Ludwig, A. Buil, A. Bensimon, M. Soste, T. D. Spector, E. T. Dermitzakis, B. C. Collins, L. Malmström and R. Aebersold, *Nat. Biotechnol.*, 2017, **35**, 781–788.

187 M. Pinkse, G. Evaristo, M. Pieterse, Y. Yu and P. Verhaert, *EuPA Open Proteomics*, 2014, **5**, 32–40.

188 V. Casanova, F. H. Sousa, P. Shakamuri, P. Svoboda, C. Buch, M. D'Acremont, M. A. Christophorou, J. Pohl, C. Stevens and P. G. Barlow, *Front. Immunol.*, 2020, **11**, 1–14.

189 S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann and J. Cox, *Proteomics*, 2015, **15**, 1453–1456.

190 P. M. Palagi, D. Walther, M. Quadroni, S. Catherinet, J. Burgess, C. G. Zimmermann-Ivol, J.-C. Sanchez, P.-A. Binz, D. F. Hochstrasser and R. D. Appel, *Proteomics*, 2005, **5**, 2381–2384.

191 N. C. Parsley, P. W. Sadecki, C. J. Hartmann and L. M. Hicks, *J. Nat. Prod.*, 2019, **82**, 2537–2543.

192 N. C. Parsley, O. L. Williams and L. M. Hicks, *J Am Soc Mass Spectrom*, 2020, In revisions.

193 J. M. Chick, D. Kolippakkam, D. P. Nusinow, B. Zhai, R. Rad, E. L. Huttlin and S. P. Gygi, *Nat. Biotechnol.*, 2015, **33**, 743–749.

194 C. DM and C. JS, *Proteomics*, 2002, **2**, 1426–1434.

195 K. Verheggen, H. Ræder, F. S. Berven, L. Martens, H. Barsnes and M. Vaudel, *Mass Spectrom. Rev.*, 2020, **39**, 292–306.

196 Z. An, L. Zhai, W. Ying, X. Qian, F. Gong, M. Tan and Y. Fu, *Mol. Cell. Proteomics*, 2019, **18**, 391–405.

197 A. Guthals, J. D. Watrous, P. C. Dorrestein and N. Bandeira, *Mol. Biosyst.*, 2012, **8**, 2535–2544.

198 J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira and P. C. Dorrestein, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E1743-52.

199 J. Y. Yang, L. M. Sanchez, C. M. Rath, X. Liu, P. D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. De Felicio, A. Fenner, W. R. Wong, R. G. Linington, L. Zhang, H. M. Debonsi, W. H. Gerwick and P. C. Dorrestein, *J. Nat. Prod.*, 2013, **76**, 1686–1699.

200 I. Verrastro, S. Pasha, K. Jensen, A. Pitt and C. Spickett, *Biomolecules*, 2015, **5**, 378–411.

201 K. Z. Hein, H. Takahashi, T. Tsumori, Y. Yasui, Y. Nanjoh, T. Toga, Z. Wu, J. Grötzinger, S. Jung, J. Wehkamp, B. O. Schroeder, J. M. Schroeder and E. Morita, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 13039–13044.

202 B. O. Schroeder, D. Ehmann, J. C. Precht, P. A. Castillo, R. Küchler, J. Berger, M. Schaller, E. F. Stange and J. Wehkamp, *Mucosal Immunol.*, 2015, **8**, 661–671.

203 M. Shabab, M. F. F. Arnold, J. Penterman, A. J. Wommack, H. T. Bocker, P. A. Price, J. S. Griffitts, E. M. Nolan and G. C. Walker, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 10157–10162.

204 K. W. Bensch, M. Raida, H.-J. Mägert, P. Schulz-Knappe and W.-G. Forssmann, *FEBS Lett.*, 1995, **368**, 331–335.

205 J. Wendler, D. Ehmann, L. Courth, B. O. Schroeder, N. P. Malek and J. Wehkamp, *Infect. Immun.*, 2018, **86**, 1–12.

206 J. Raschig, D. Mailänder-Sánchez, A. Berscheid, J. Berger, A. A. Strömstedt, L. F. Courth, N. P. Malek, H. Brötz-Oesterhelt and J. Wehkamp, *PLoS Pathog.*, 2017, **13**, 1–18.

207 G. Paone, A. Wada, L. A. Stevens, A. Matin, T. Hirayama, R. L. Levine and J. Moss, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8231–8235.

208 A. Bonucci, E. Balducci, M. Martinelli and R. Pogni, *Biophys. Chem.*, 2014, **190–191**, 32–40.

209 E. Kudryashova, S. M. Seveau and D. S. Kudryashov, *Biol. Chem.*, 2017, **398**, 1069–1085.

210 Q. Li, M. R. Shortreed, C. D. Wenger, B. L. Frey, L. V. Schaffer, M. Scalf and L. M. Smith, *J. Proteome Res.*, 2017, **16**, 1383–1390.

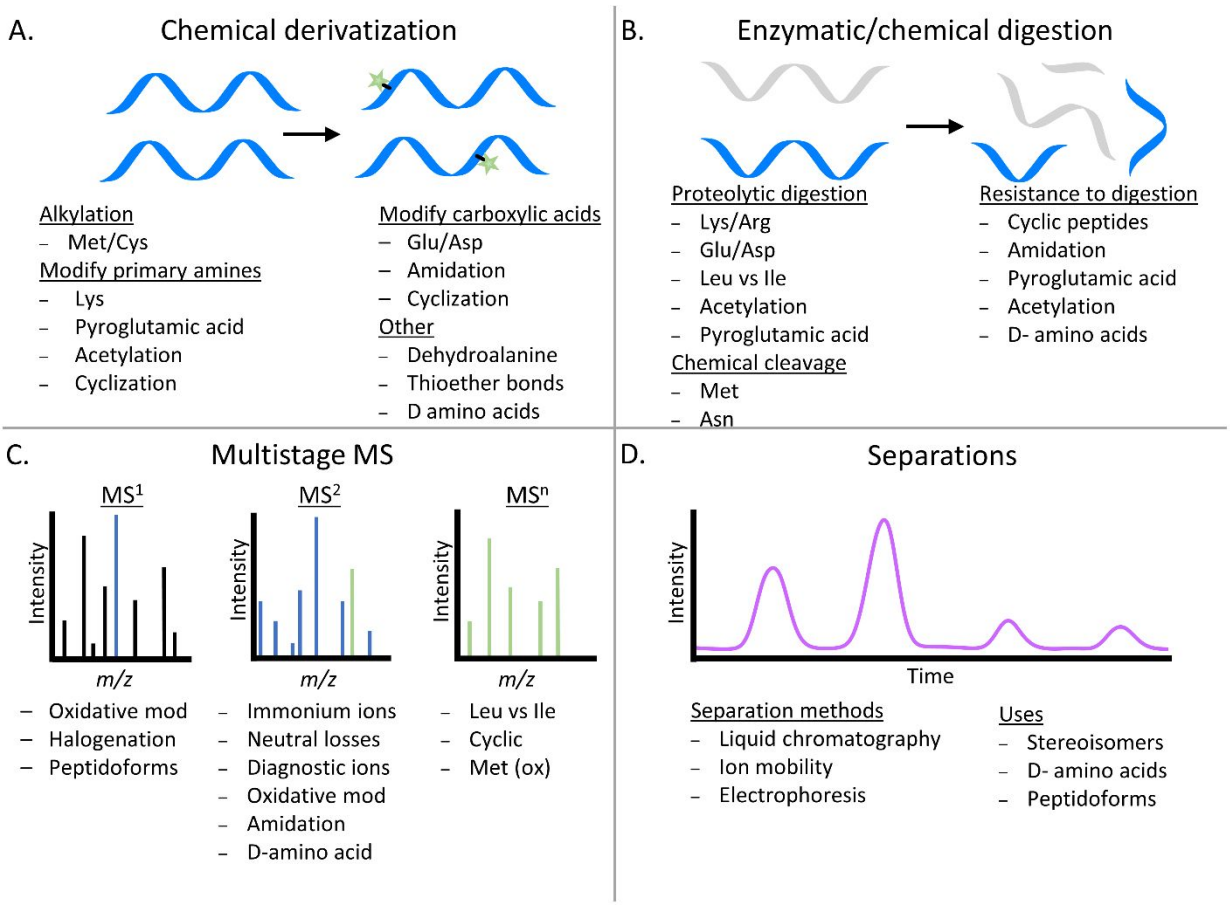


Figure 1. Strategies to identify amino acid composition and post-translational modifications contain four main categories: (A) chemical derivatization, (B) enzymatic/chemical cleavage, (C) multistage mass spectrometry, and (D) separations.

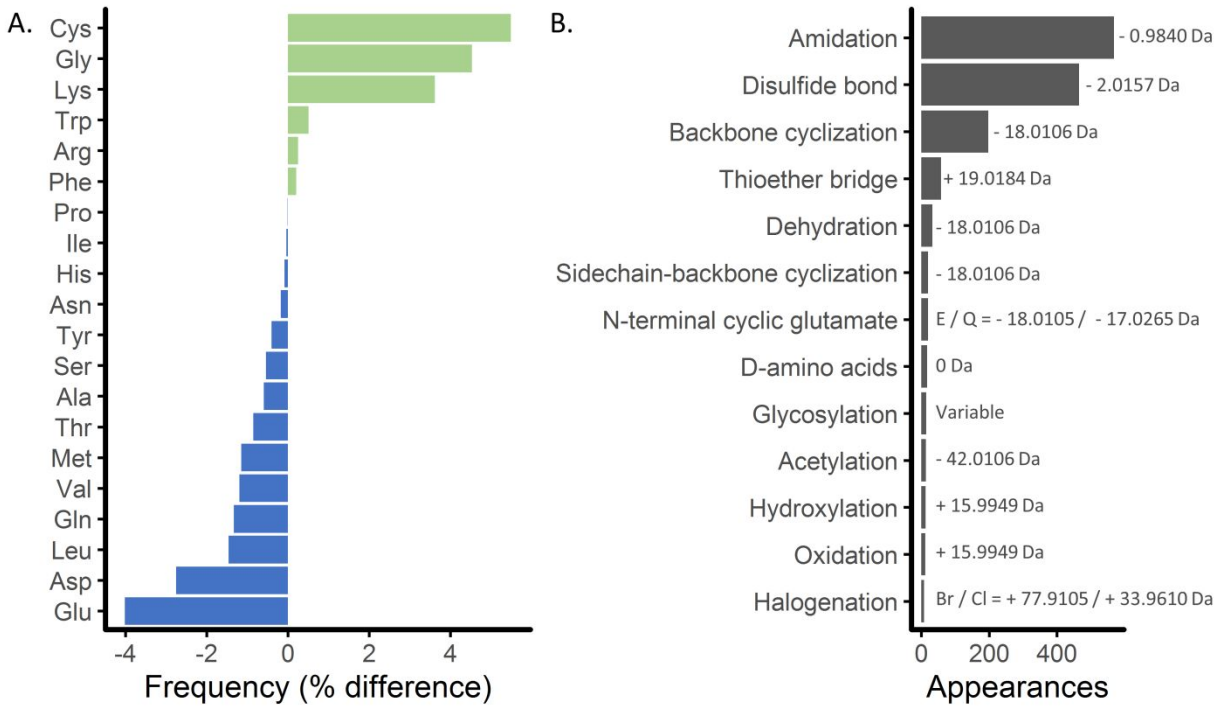


Figure 2. The Antimicrobial Peptide Database (APD) contains 3076 natural, ribosomally synthesized AMPs. (A) Amino acid residue frequency differs between these AMPs and the SwissProt database of annotated proteins. (B) The top fourteen AMP PTMs vary widely in frequency and mass shift. The mass shift of glycosylation is noted as variable because a wide variety of glycan groups can be added. Data was retrieved from the APD on February 20, 2020 and SwissProt database on June 6, 2020.

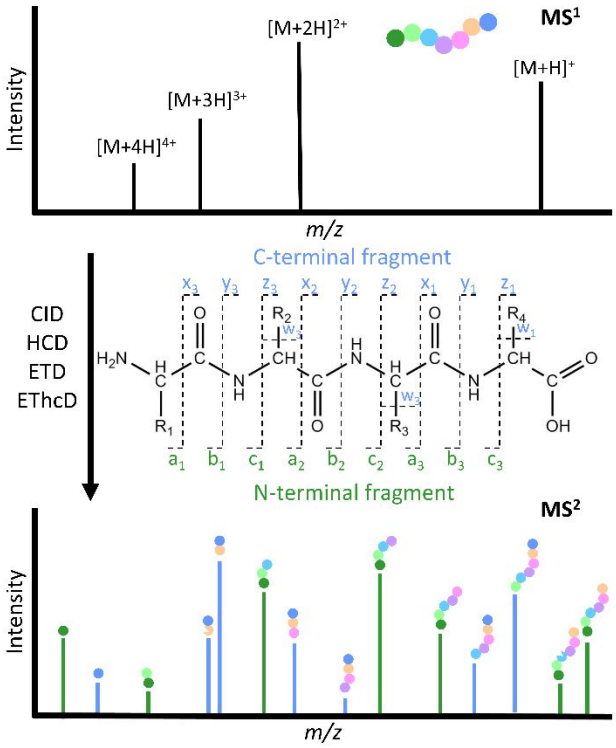


Figure 3. Mass spectrometry is a powerful tool for peptide sequencing. Peptides are identified by their intact mass in MS^1 prior to fragmentation. Methods for MS^2 fragmentation include collision induced dissociation (CID), higher-energy C-trap dissociation (HCD), electron transfer dissociation (ETD), and electron-transfer/higher-energy collision dissociation (EThcD). Peptides fragment predictably along the peptide backbone producing fragments which contain the N- (green; *a*-, *b*-, *c*-) or C- (blue; *x*-, *y*-, *z*-) terminus. Other fragments resulting from the partial cleavage of side-chains, such as *w*- ions (*z*- ion with side-chain loss, noted on fragmentation map) or neutral losses (represented by partial circle), can also provide sequence information. Fragmentation methods produce different types of backbone ions and may produce complementary information. Intact masses and fragmentation patterns are used to determine peptide sequences via *de novo* sequencing and/or database searching.

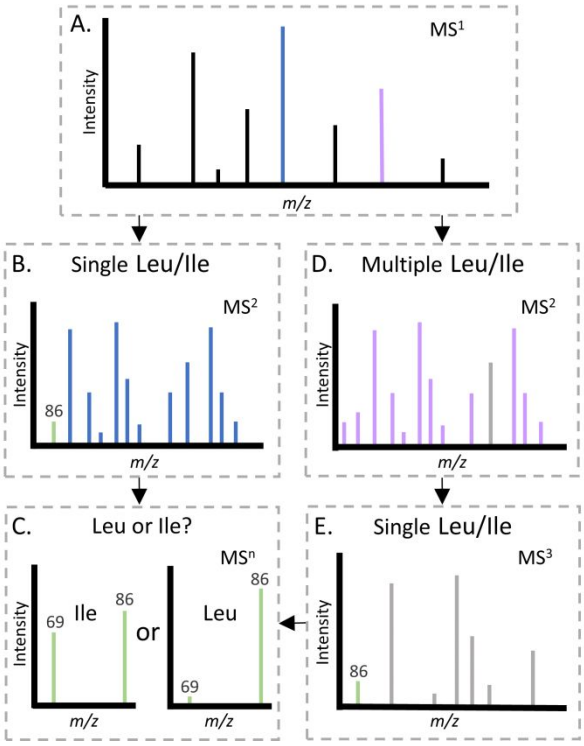
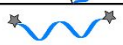


Figure 4. Multistage mass spectrometry can be used to differentiate structural isomers Leu and Ile. If the peptide of interest (A, blue) contains only a single Leu/Ile, (B) it is selected for fragmentation and produced (green) an Leu/Ile immonium ion (86.0961 *m/z*). (C) If the peptide contains Ile, the immonium ion will produce a 69 *m/z* ion that is greater than ten percent of the 86 *m/z* precursor upon additional fragmentation. The peptide contains a Leu residue if the 69 *m/z* ion is unstable and less than ten percent of the precursor. In cases where (A, purple) the peptide of interest contains multiple Leu/Ile residues, (D) it is selected for fragmentation and produces (grey) an MS^3 product ion containing a single Leu/Ile. (E) This MS^3 product ion is additionally fragmented yielding (green) an immonium ion from a single Leu/Ile. (C) The simplified immonium ion is fragmented to identify residue. This process is repeated until all Leu/Ile are differentiated.

Modification	Example	Primary amine derivatization	Carboxylic acid derivatization	General aminopeptidase	Specific aminopeptidase	Carboxypeptidase	Neutral Loss
Unmodified		✓	✓	✓	⊘	✓	⊘
N-term modification		⊘	✓	⊘	✓	✓	⊘
C-term modification		✓	⊘	✓	⊘	⊘	✓
N-term cyclization		⊘	✓	⊘	✓	⊘	⊘
C-term cyclization		✓	⊘	✓	⊘	✓	⊘
N- and C-term modification		⊘	⊘	⊘	✓	⊘	✓
C-term cyclization N-term modification		⊘	⊘	⊘	✓	⊘	⊘
N-term cyclization C-term modification		⊘	⊘	⊘	⊘	⊘	✓
N-to-C cyclization		⊘	⊘	⊘	⊘	⊘	⊘
D-amino acid near N-terminus		✓	✓	⊘	✓	✓	✓

✓ Positive result

⊘ Negative result

Figure 5. N- and C- terminal modifications can be difficult to differentiate within a single experiment, but a series of derivatizations, digestions, and fragmentation experiments can be used to facilitate characterization. Successful results (e.g. digestion, derivatization, neutral loss detection) provides direct evidence. Negative results, however, are equally insightful because they indicate that the peptide did not contain the necessary conditions for successful transformation. Taken together, the results from multiple experiments can facilitate termini characterization.

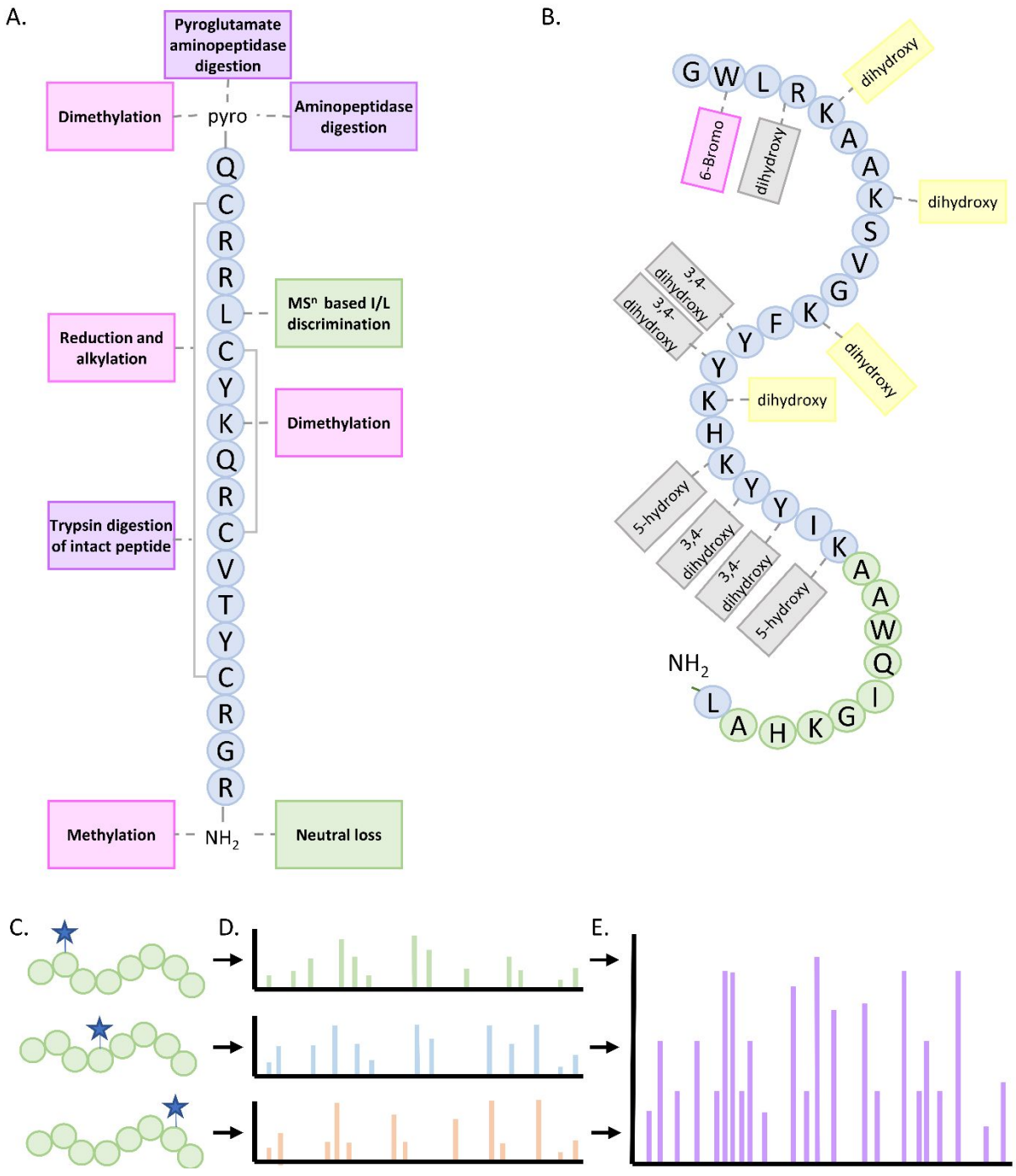


Figure 6. The antimicrobial peptides gomesin and styelin D are used as case studies to examine the benefits and challenges of MS-based characterization methods. (A) An alternative theoretical workflow is proposed for gomesin which uses chemical derivatization (pink), multistage MS (green) and enzymatic cleavage (purple) to identify key sequence features and facilitate *de novo* sequencing, thus providing a feasible route to characterization without peptide isolation or genomic information. (B) Variable lysine hydroxylations (yellow) and bromination (pink) result in complex fragmentations spectra which make the unusual and complex AMP styelin D difficult to manually sequence. However, the unmodified residues near the C-terminus (green) could be manually sequenced from the MS/MS spectra for the intact peptide, used as a

sequence tag, and combined with genomic information to identify the primary sequence of styelin D. (C) Peptides which vary only in PTM localization have the same exact mass but (D) have distinct MS/MS spectra. (E) If same mass peptides are co-isolated for fragmentation, they produce far more complicated and difficult to interpret chimeric spectra.

Table 1. Chemical derivatizations. Iodoacetamide (IAM); Methyl methanethiosulfonate (MMTS); N-ethylmaleimide (NEM); Aminobutyric acid (Abu).

Derivatization		Residues	Mass shift (Da)	Notes	Section	Ref
Alkylation	IAM	Cys	Reduced: + 57.0215 Oxidized: + 58.0293	For Met - control pH (< 3) and use long reaction times (>48 h)	2.1, 2.2	38-40
		Met	+ 58.0293			
	MMTS	Cys	Reduced: + 45.9880 Oxidized: + 46.9958			
	NEM		Reduced: + 126.0555 Oxidized: + 127.0633			
Dimethylation		Lys, N-term	+ 28.0313	Also modifies free N-terminus	2.3, 3.1, 3.3	41
Methyl esterification		Glu, Asp, C-term	+ 14.0156	Also modifies free C-terminus	2.3, 3.1, 3.3	42
Deuterated reductive desulfurization		Dehydrated residue	+ 4.0282	2 D incorporated into Ala/Abu	3.2	43
		Thioether bridge	- 27.9439	1 D incorporated into Ala/Abu		

Table 2. Proteases and chemical cleavage agents.

Cleavage agent	Specificity	Notes	Section	Ref
CNBr	Met	Greater specificity in 70% formic acid Nonspecific cleavage: Tyr Reduced cleavage: MetSer and Met-Thr	2.2	44
Trypsin	Lys, Arg	pH 7.5	2.3	45
Arg-C	Arg	pH 7-8	2.3	45
Lys-N	Lys	pH 9.5	2.3	45
Lys-C	Lys	pH 7-9	2.3	45
Glu-C	Glu	Ammonium buffers, pH 4-9	2.3	46
	Asp and Glu	Phosphate buffers, pH 4-9		
Asp-N	Asp and cysteic acid	pH 4-9	2.3	45
Chymotrypsin	Tyr, Phe, Trp, Leu	pH 7.8-8	2.4	45
Thermolysin	Leu, Phe, Val, Met	pH ~ 7.5, 65 ^{o}C	2.4	47
Pepsin	Phe, Leu, Tyr, Trp	pH 1-4	2.4	45
Diaceroxyiodobenzene	Asn	N-terminus must be blocked	2.5	49
Carboxypeptidase Y	C-terminal amino acids	Blocked by Pro, hydroxyPro, Arg	2.3, 3.1, 3.3	49
Aminopeptidase M	N-terminal amino acids	Blocked by N-terminal Asp, Glu, Pro, D-amino acid or modified N-termini	2.3, 3.1, 3.3, 3.7	50
Pyroglutamate aminopeptidase	N-terminal pyroglutamic acid	Sensitive to urea	3.1	51
Acylamino acid releasing enzyme	N- terminal acetylated residues	Inhibited by many buffers	3.1	51
Peptide N-Glycosidase F	N-linked oligosaccaharides	Ammonium bicarbonate, pH 8	3.5	52
Endoglycosidase H	Limited number of N-linked oligosaccharides	Ammonium acetate, pH 5.5	3.5	52
Hydrazinolysis	Unreduced O- and N-linked oligosaccharides	Release N- and O-: 95 ^{o}C, 5 h Release O-: 60 ^{o}C, 4 h	3.5	53

Table 3. Feature specific ions and losses. Asterisks indicate immonium ions that can be differentiated from interfering ions with additional stages of fragmentation. Related ions are non-immonium ion internal fragments.

Features	Immonium ion	Immonium ion fragments	Related ions	Neutral Loss	Ref.
Ala	44.0491	-	-	-	54
Arg	129.1131	-	59, 70, 73, 78, 100, 112	17.0266, 34.0531, 43.0296, 59.0484, 60.0211, 62.0242, 101.0953	54-57
Asp	88.0381	-	-	18.0106, 27.9949, 46.0055, 60.0211	54, 55, 57
Asn	87.0541	-	70	17.0266, 45.0215, 59.0371	54-57
Cys	76.0211	-	-	75.0558, 90.0013, 107.0279	54, 55, 57
Gly	30.0331	-	-	-	54
Gln	101.0701	-	84, 129	17.02655, 45.0215, 59.0371, 71.0371	54-57
Glu	102.0541	-	-	18.0106, 27.9949, 46.0055, 76.0398, 89.0477	54, 55, 57
His	110.0701	-	82, 121, 123, 138, 166	82.0531, 83.0609, 84.0688	54-57
Ile	86.0961*	69.0970	72	46.0657	54, 56, 58
Leu	86.0961*	-	72	56.0626, 60.0813, 73.0891	54, 55
Lys	101.1061	-	70, 84, 112, 129	17.0266	54
Met	104.0521	-	61	91.0456	54, 55
Phe	120.0801	-	91	-	54
Pro	70.0621	-	-	-	54
Ser	60.0421	-	-	18.0106, 35.0371	54, 55
Thr	74.0591	-	-	18.0106, 35.0371	54, 55
Trp	159.0911	-	117, 130, 170, 171	-	54
Tyr	136.0751	-	91, 107	108.0575	54, 55
Val	72.0801	-	-	-	54
Carbamidomethyl Met	-	-	-	105.0248	59
N-terminal amidation	-	-	-	17.0266	60
Pyroglutamic acid	58.9946	-	171.0674	-	54
Oxidized Met	117.0650	-	-	63.9983	62
5-hydroxy Trp	175.0860*	158.1, 148.1, 146.1	-	-	61
2-hydroxy Trp	175.0860*	158.1, 130.1	-	-	61
3-hydroxy Tyr	152.0700*	135.1, 107.1	152.0700	-	61

1