nature reviews methods primers

https://doi.org/10.1038/s43586-023-00205-2

Primer

Check for updates

Peptidomics

Roland Hellinger[1,5], Arnar Sigurdsson[2,5], Wenxin Wu[3,5], Elena V. Romanova[4,5], Lingjun Li[3] ✉, Jonathan V. Sweedler[4] ✉, Roderich D. Süssmuth[2] ✉ & Christian W. Gruber[1] ✉

Abstract

Peptides are biopolymers, typically consisting of 2–50 amino acids. They are biologically produced by the cellular ribosomal machinery or by non-ribosomal enzymes and, sometimes, other dedicated ligases. Peptides are arranged as linear chains or cycles, and include post-translational modifications, unusual amino acids and stabilizing motifs. Their structure and molecular size render them a unique chemical space, between small molecules and larger proteins. Peptides have important physiological functions as intrinsic signalling molecules, such as neuropeptides and peptide hormones, for cellular or interspecies communication, as toxins to catch prey or as defence molecules to fend off enemies and microorganisms. Clinically, they are gaining popularity as biomarkers or innovative therapeutics; to date there are more than 60 peptide drugs approved and more than 150 in clinical development. The emerging field of peptidomics comprises the comprehensive qualitative and quantitative analysis of the suite of peptides in a biological sample (endogenously produced, or exogenously administered as drugs). Peptidomics employs techniques of genomics, modern proteomics, state-of-the-art analytical chemistry and innovative computational biology, with a specialized set of tools. The complex biological matrices and often low abundance of analytes typically examined in peptidomics experiments require optimized sample preparation and isolation, including in silico analysis. This Primer covers the combination of techniques and workflows needed for peptide discovery and characterization and provides an overview of various biological and clinical applications of peptidomics.

Sections

[1]Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria. [2]Institut für Chemie, Technische Universität Berlin, Berlin, Germany. [3]School of Pharmacy and Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA. [4]Department of Chemistry, University of Illinois, Urbana, IL, USA. [5]These authors contributed equally: Roland Hellinger, Arnar Sigurdsson, Wenxin Wu, Elena V. Romanova. ✉e-mail: lingjun.li@wisc.edu; jsweedle@illinois.edu; suessmuth@chem.tu-berlin.de; christian.w.gruber@meduniwien.ac.at

Introduction

Peptides occur in all three domains of life. Their native functions comprise acting as peptide hormones for cellular signalling, as secretory peptides for interspecies communication and interaction, as predatory peptide toxins and as defence peptides against microorganisms, viruses and herbivores[1]. Peptides are important molecules that play a fundamental role in human physiology and pathology. In addition, natural sources including animals, plants, fungi and microorganisms provide rich sources of biologically active peptides[1,2]. Research has focused on investigating and understanding the biological function of peptides and their potential as disease biomarkers or therapeutic lead compounds and drugs[3,4].

Peptides are molecular entities of amino acids linked via amide bonds rendering the peptide backbone or peptide chain. Although the size discrimination between a large peptide and a small protein is rather arbitrary, the scientific community typically refers to a peptide containing between 2 and 50 amino acids. Most peptides are encoded by DNA or RNA and, hence, are products of translation, transcription and the cellular protein manufacturing machinery. Typically, they contain many and often complex post-translational modifications, including side-chain modifications, disulfide bond formation, individual residue isomerization and, sometimes, head-to-tail or side-chain cyclizations. Non-ribosomal enzymatically synthesized peptides (typically produced by bacteria and fungi) further increase the variety of peptide species by incorporating non-proteinogenic building blocks and a set of additional tailoring reactions. Overall, naturally occurring peptides range between simple linear and often heavily modified molecules, which occupy a large and unique chemical space in between small organic molecules and larger proteins[5].

The highly diverse and easily modified peptides from nature are among the most interesting molecules. Peptides can be obtained from various sources and can range from structurally simple to highly functionalized and complex structures, giving them a plethora of biological features (Fig. 1). Although differences in the extent of modification might appear extreme, they could still exert the same biological function. Owing to their relatively small size and chemical potential, peptides can possess valuable chemical information for the design of novel drugs[6].

The term peptidomics was introduced to define a strategy for the direct measurement and structural characterization of endogenous peptides in biological systems in a high-throughput manner, with robust and unprecedented sensitivity[7–9]. Since then, peptidomics became a fast developing and progressing multidisciplinary field that combines state-of-the-art separation techniques including liquid chromatography, modern mass spectrometry technologies, innovative bioinformatics and statistics for qualitative and quantitative analysis of peptides relevant to fundamental biology and human health sciences. Although the field of peptidomics resorts to various technologies that have been developed for proteomics, analytical chemistry and genomics, it is the unique sample preparation and specialized combination of these tools and methods that make peptidomics a unique discipline and research field.

Peptidomics refers to a system-level study of a set of analytes with the aim to describe the number and identity as well as the relative or absolute levels of peptides. As with other biological omics fields, peptidomics has a significant overlap with genomics and proteomics. Bioinformatic approaches have been utilized to discover genes encoding for peptides and proteins the peptide product could have originated from, for potential use as novel medicinal products or as clinical biomarkers. With more newly discovered peptides, as the algorithms used become more powerful, bioinformatics and high-resolution mass spectrometry are now shaping the future of biological sciences.

Peptides can have various applications, ranging from food preservatives to therapeutic agents in humans (Fig. 1). However, the use of peptides goes even beyond that, as peptides can act as catalysts, improving the yields and enantioselectivity of chemical synthesis[10]. Highly efficient and selective catalysis was originally considered to be the domain of proteins, but peptides can be modified to catalyse simple chemical reactions. Biomaterials are now a trending topic, where sustainability and recyclability will be a requirement for future industries[11,12]. Peptides are accessible and chemically diverse oligomers, making them a popular choice for research and development, with ample application outside biological sciences. However, the discovery and characterization of new peptides lays the foundation for their future utilization.

This Primer aims to define and describe peptidomics technologies, tools and workflows for identification and analysis of endogenous peptides in a qualitative and quantitative manner. The field has rapidly advanced in the past three decades with the development of computational and mass spectrometry-based techniques. Remaining analytical challenges and pending questions in peptidome analysis are discussed for selected applications. The Primer covers various in silico and peptide analysis methodologies including de novo sequencing, high-throughput and automated peptidomics workflows, peptide imaging and quantitative mass spectrometry approaches. The application of peptidomics in biology, in drug discovery and in the clinic to identify novel biomarkers and understand disease mechanisms are discussed. Lastly, reproducibility and data deposition are discussed, followed by current limitations and the outlook regarding ongoing developments in the field of peptidomics.

Experimentation

Modern peptidomics workflows encompass the analysis of genetic information, characterization of peptides and computational processing of the data. Although this multilevel analysis allows for more comprehensive read-outs, a single-step analysis also qualifies to be counted into the field of peptidomics. Genetic information is accessed by peptide precursor mining or metabolic network analysis. Working at the peptide level can be subdivided into several steps, including sample preparation and clean-up, mass spectrometric analysis and data evaluation or integration.

Sample preparation

Multiple workflows have been described that proved useful for analysis of various peptide entities. State-of-the-art peptidomics workflows are amenable to complex samples, as they are obtained from plant, microbial cell or animal/human tissue extractions. However, sample processing steps are usually implemented to enrich analytes over matrix compounds and to concentrate low-abundance analytes. This Primer describes common pipelines; several of these methods can be modularly combined for workflows suitable to address the needs of any specific research questions. Approaches are optimized by trial and error to obtain the desired outcomes.

Sample harvest, cell lysis and extraction. Depending on the source of peptides, there are several harvest and extraction procedures available. A common problem during sample preparation is the degradation of peptides – especially low-abundance peptides, for example by cellular proteases. For instance, neuropeptides and peptide hormones are biosynthesized as larger precursor proteins; they are converted into

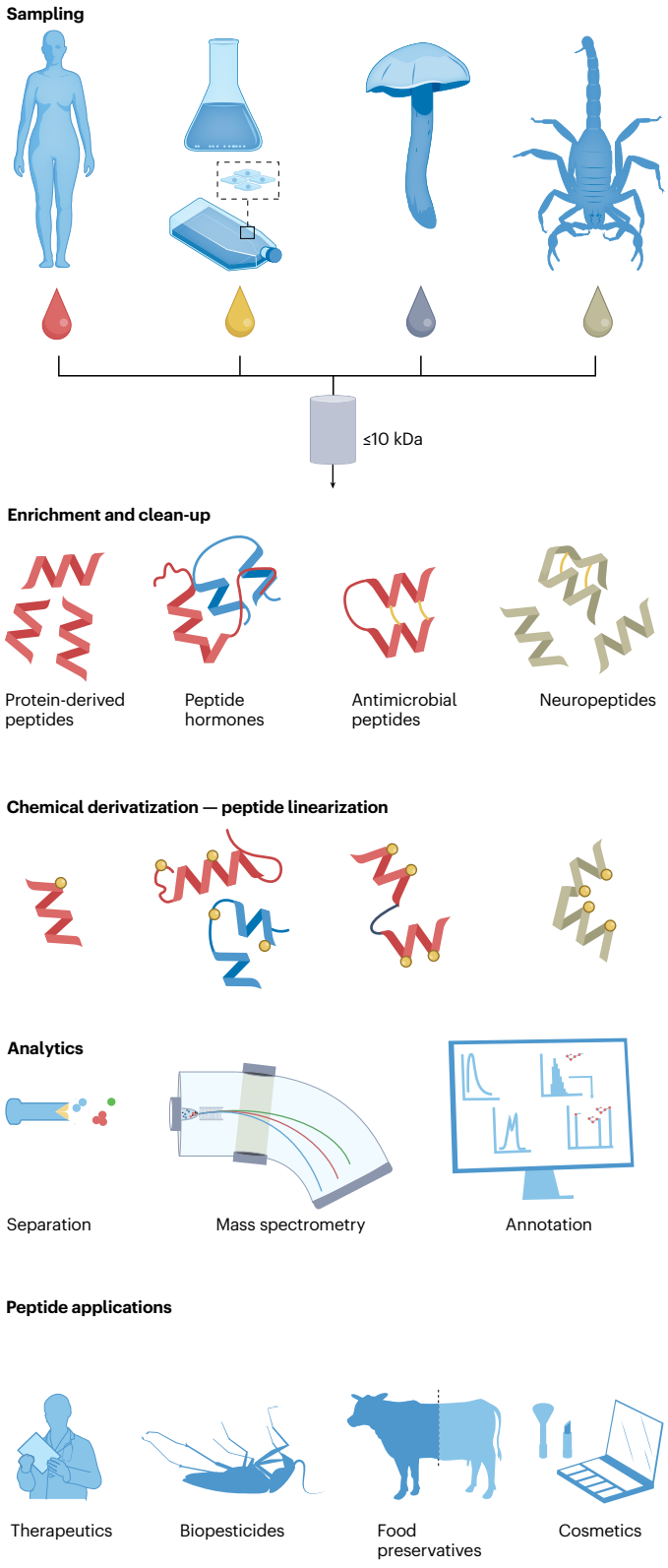


Fig. 1 | Peptides, from analysis to application. The sources of peptides are vast, and many serve diverse biological roles. However, peptides are often difficult to detect without proper enrichment or clean-up strategies for use in downstream applications. Subsequent modification or derivatization of the peptides may be required to elucidate their sequence or structural features, which is essential for further design or use of discovered peptides in medical or industrial applications.

their bioactive form by processing enzymes that cleave their precursor into several final active messengers or hormones[13]. As a neuropeptide's biological message is transient, active peptides become inactivated by proteases[14]. Optimizing sampling approaches is important to counteract rapid postmortem/post-harvest degradation. To increase the yield of active peptides for analysis, degrading enzymes can be inactivated by heating tissue/cells, or by addition of protease inhibitors or chaotropic agents[15,16]. Extended heat treatments should be avoided as peptides are prone to chemical reactions, such as hydrolysis and oxidation. Processing/storing samples at low temperature helps avoid peptide degradation/modification. These steps are less crucial for stabilized peptides, such as plant defence peptides or venom-derived peptide toxins. For efficient extraction of peptide analytes, the cells/tissue must be broken up by cell lysis. For instance, cells derived from cell cultures or from collagen-treated tissues can be lysed with hypotonic buffers followed by ultra-sonication to disintegrate the phospholipid bilayers to release intracellular material[17]. Throughout these steps, a buffer amenable to dissolving the peptides should be used, which can be aqueous or organic depending on the nature of the sample. Obviously, the heat treatment of biological samples for protease inactivation will already be the first step to induce lysis. The lysis of microorganisms or plants requires specialized protocols[18]. For instance, to extract peptides from plants, the source material is mechanically ground or pulverized (using a grinder/shredder or liquid N_2 and mortar/pestle); the peptides are then extracted by use of organic solvents in combination with alcohols/aqueous buffers[19]. The organic solvent used to remove chlorophyll not only helps break down cellular membranes to release the peptides into the buffer but also inhibits any unwanted protease degradation. Generally, detergent-based lysis using cationic, zwitterionic or anionic detergents is not compatible with mass spectrometry (but there are alternative methods available[20]). The efficient and repeatable peptide extraction from biological matrices is key for further sample processing, and to achieve maximum concentration of the peptide analytes.

Enrichment and clean-up strategies. After extraction, crude biological samples contain a low abundance of peptides, as they also contain salts, lipids, proteins and carbohydrates[21], which makes purification steps necessary. The complexity of this matrix background (molecules other than the peptides of interest) can impair the ability of mass spectrometry to identify the peptides of interest. Although less complicated sample preparation techniques are needed for peptidomics compared with proteomics, several clean-up techniques are effective for enriching the peptides of interest prior to instrumental analysis[22]. Solid-phase extraction (SPE) is commonly applied for sample concentration and desalting, as a rapid, stand-alone tool. This is especially important if the samples are being analysed directly by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry, a soft ionization technique, as high salt concentrations suppress analyte–matrix crystallization and ionization. For reversed-phase SPE there are many different types of cartridges (ranging in size, volume, chemistry, vendor) available. They can be used for single or multi-channel sample processing using vacuum manifold systems, which can handle up to 24 cartridges simultaneously, or in 96/384-well format, utilizing SPE containing plates for automated sample clean-up. For rapid clean-up of small

sample volumes, ZipTips or alternative pipette tip SPE devices have become useful[23]. Alternative but low-resolution (and sometimes low-recovery) methods for sample clean-up are molecular weight cut-off or nanofiltration devices and classical sample precipitation techniques. A more efficient method for sample clean-up is liquid chromatography. This technique can be used offline or coupled directly to mass spectrometry. It is compatible with numerous separation chemistries. One example is food-derived peptides, a large class of bioactive molecules usually ranging in size from 2 to 20 amino acids. They are generated from enzymatic food hydrolysates during digestion or fermentation. The isolation of peptides from the hydrolysate matrix is achieved by a combination of ultrafiltration and chromatographic separation techniques, such as reversed-phase, ion exchange and size exclusion chromatography[24]. The isolated peptides are further analysed to the amino acid level using amino acid quantitation, Edman degradation or mass spectrometry-based de novo sequencing.

Chemical derivatization and peptide labelling. For certain peptidomics applications, such as discovery, isolation and analysis of cysteine-rich peptides, chemical derivatization of cysteine residues may be beneficial. Disulfide bridged or cysteine knot peptides are chemically reduced to the sulfhydryl-containing compound with mild reduction reagents and, then, immediately converted into stable derivatives. Thiol-containing reductants are β-mercaptoethanol, glutathione and dithiothreitol (DTT) supporting sulfhydryl deprotonation under basic conditions. Phosphine-based reagents, for example tris(2-carboxyethyl)phosphine (TCEP), are considered more stable than thiols and are amenable to all pH conditions[25]. The cysteine sulfhydryl group derivatization makes further use of a few standard reagents, preferably halogenated acetate derivatives (for example, iodoacetamide, iodoacetic acid), N-substituted maleimides (for example, *N*-ethyl maleimide), 4-vinylpyridine or thiosulfonate reagents (for example, methyl-methane thiosulfonate)[26,27]. Iodoacetamide has become a standard working horse for proteomics and peptidomics applications. Derivatization was also explored for peptide quantification via cysteine derivatization. For instance, owing to its chromophore, 5,5-dithiobis(2-nitrobenzoate) (DTNB) provides labelled peptides that can be quantitatively analysed by HPLC–UV detectors. Other labelling approaches established in the peptidomics field support the detection and quantitation or comparative analysis of peptidome samples.

Labelling of peptides with stable isotopes is a common strategy in peptidomics to increase sensitivity and detection and to enable quantitative analysis. Labelling of peptides of interest with stable isotopes in biological models can be achieved by metabolic, enzymatic and chemical strategies. Metabolic labelling introduces stable isotopes through metabolic incorporation in vivo, often via diet or culture media, and thus leads to long experimental procedures. Several chemistry-based labelling protocols have been implemented that enable stable isotope tagging of peptides in vitro[28–32]. There are several commercial tags available, for example tandem mass tags (TMT)[33] and isobaric tags for relative and absolute quantitation (iTRAQ)[34]; there are also many low-cost alternatives available, including *N*,*N*-dimethyl leucine (DiLeu)[31], deuterium isobaric amine reactive tag (DiART)[35] and 10-plex isobaric tags (IBT)[36].

Peptide separation

Crude cell or tissue extracts typically present high chemical complexity and a large concentration range of diverse compounds. As peptides may exist in solution as charged molecules with differing degrees of hydrophobicity, they are amenable to several separation technologies. Although reversed-phase liquid chromatography has been widely used, size exclusion, ion exchange and mixed-mode applications are gaining attraction with continuous development of new sorbent materials[37]. Chromatographic separation systems are either directly hyphenated to mass spectrometry via the electrospray ionization (ESI) interface or performed offline[38]. Reversed-phase chromatography promises high peak capacity and excellent resolution. Particles in use are modified silica gels with alkyl group substitution, with pore sizes typically ranging from 110 to 300 Å to facilitate adsorptive analyte–stationary phase interactions. The alkyl chain length is the major determinant for overall hydrophobicity of the material and the retention power for analytes. Octadecyl (C_{18}), octyl (C_8) or butyl (C_4) alkylated silica materials are the most common. The selectivity for analytes and their peak shape can further be improved by end-capping, cross-linking or other modifications. For instance, trimethylsilyl or polar groups are used for fine-tuning selectivity, resolution and retention capacity of the stationary phase. The optimal choice of reversed-phase stationary material depends on the structural and chemical diversity of peptide analytes. For example, C_8 material commonly enables better separation of basic and neutral molecules (for example, tryptic peptides) under acidic conditions. The mobile phases in reversed-phase applications are aqueous and organic solvents, usually methanol or acetonitrile with acidic modifiers. Trifluoroacetic acid (TFA) is a strong ion-pairing counter-ion providing very reliable and reproducible peak shapes and overall separations[39]. TFA is unfortunately not well compatible with the ESI technique commonly applied in liquid chromatography–mass spectrometry (LC-MS) systems. Here, formic acid is used instead and ammonium formate or ammonium acetate can be used as the mass spectrometry-compatible buffer system.

More specialized applications such as affinity-based and immobilized-metal affinity chromatography are used to enrich or isolate phosphopeptides from tissue extracts, whereas hydrophilic liquid interaction chromatography is utilized for separation of glycopeptides[40]. For mass and volume-limited samples, capillary electrophoresis can be used as a front-end separation method. One advantage of capillary electrophoresis is that it is compatible with one to two orders of magnitude smaller sample volumes than liquid chromatography systems, and therefore is the best choice for cellular and subcellular peptidomics.

The most recent addition to peptide separation approaches is ion mobility spectrometry (IMS) that sorts and separates gas-phase ions according to their 3D shapes. Placing an IMS module between the source and the mass analyser increases ion utilization efficiency, improves sensitivity and specificity of detection, and broadens the dynamic range. Because of the introduction of a new range of IMS instruments, this approach is rapidly gaining application in peptide characterization and quantitation.

Mass spectrometry technologies

Mass spectrometry instrumentation. There are various different mass spectrometry systems available, which can be categorized by the ionization technique, resolution of the mass analyser, single or tandem set-up of the system or mass analyser type. The main soft ionization techniques are MALDI[41,42] and ESI[38]. Fourier transform ion cyclotron resonance mass spectrometers (FTICR-MS) are high-resolution instruments suitable for peptide mapping and characterization via accurate mass matching and mass spectrometry imaging (MSI) studies. High-speed mass spectrometry systems, typically equipped with a quadrupole,

time-of-flight (TOF) or ion trap mass analyser, are useful for quantitative analysis[43,44]. Most of these systems have the capacity to perform tandem mass spectrometry and fragmentation experiments in the MS/MS or MS/MS/MS mode (such as ion trap and orbitrap systems)[45,46]. Peptide fragmentation uses different techniques, such as post-source decay, high-energy collision-induced dissociation or electron transfer dissociation methods[47–49]. Peptidomics studies utilize several mass analysers of various types, such as TOF, quadrupole, ion trap and, more recently due to its high resolving power, orbitrap[23] or ion cyclotron resonance analysers.

Tandem mass spectrometry workflows. Shotgun peptidomics is based on high-throughput automated sequencing and identification of endogenous peptides representative of a biological sample. Sequencing is realized by implementation of tandem mass spectrometry (MS/MS)[45] with two main protocols: data-dependent and data-independent acquisition (DDA and DIA, respectively). The benefit of DDA is high-quality MS/MS spectra resulting from a user-specified number of the most intense precursor ions in a given chromatographic time frame subjected to fragmentation. Identification of low-abundance ions is facilitated by dynamic exclusion of previously sampled precursor ions from MS/MS. Given the limited number of precursors sampled within a duty cycle, the resolution of hyphenated front-end separation platforms plays a crucial role in the complexity of the mass spectrometry spectrum and, respectively, in the degree of sample characterization via MS/MS. In DIA-based analysis, all precursor ions from the mass spectrometry survey scan are selected for MS/MS via stepping broad isolation windows across the entire *m/z* range. Implementation of DIA improves reproducibility across samples, which in turn reduces missing values and greatly improves the quantitative accuracy of peptidomics assay. Although useful for quantitation of peptides, the main limitation of DIA is dependence on reference spectral libraries, typically generated via DDA analysis of additional samples. Current efforts are channelled towards development of library-free DIA approaches[50–52].

Mass spectrometry imaging. The development of MALDI MSI maps the spatial locations and distribution patterns of the biomolecules in tissue samples[53]. Currently, MALDI MSI remains the most common method for spatial mapping of lipids, metabolites and peptides/proteins. Other than MALDI MSI, secondary ion mass spectrometry or SIMS-MSI[54], desorption ESI MSI[55] and scanning microprobe MALDI MSI[56], surface-assisted laser desorption/ionization mass spectrometry[57] and nanostructure imaging mass spectrometry[58] are also used to examine the localization of proteins/peptides. Advantages and disadvantages of the different ionization techniques have been reviewed elsewhere[59]. Unlike immunohistochemistry, radio-immunoassays and fluorescence microscopy that require extensive sample preparation and prior knowledge of the target analytes[60–62], MSI involves relatively simple sample preparation and enables the localization of hundreds to thousands of different analytes on a tissue slice in a single experiment[63]. To localize the peptide, the tissue of interest needs to be properly prepared. After dissection it is usually fresh frozen or embedded in gelatine, sodium carboxymethyl cellulose or paraffin. Formalin-fixed paraffin-embedding is a special embedding technique that can preserve the specimen under room temperature for more than a decade[64,65]. Although optimal cutting temperature (OCT) compound is a common tissue-embedding solution, the high concentration of polyethylene glycol (PEG) in the OCT compound can affect the analyte signals and, thus, OCT compound is not recommended for tissue embedding in MSI experiments[66]. Cleaning procedures to remove interferring embedding substances are needed before using OCT compound or formalin-fixed paraffin-embedded tissue for MSI experiments[67–69]. For endogenous peptide MSI experiments, sample preparation requires fewer steps; before imaging, fresh frozen or embedded tissues are sectioned into 5–20 µm slices[70]. Before imaging, background salts and lipids need to be removed from the tissue for which various washing techniques such as organic solvents have been evaluated[71]. Washing is also adaptable for peptide imaging experiments, but may increase the loss of hydrophilic peptides[72] and cause analyte diffusion[73]. Optimization of the washing procedure for formalin-fixed paraffin-embedded tissue sections shows signal enhancement for neuropeptide imaging[74]. Tissue slices with or without a washing procedure can then be analysed by imaging instruments (excluding MALDI MSI). For MALDI MSI, matrix application is needed as the last step before instrumental analysis. The matrix choice, concentration and application method are important for signal intensity and resolution and need to be carefully selected[75,76]. The resulting images from the MSI experiment can be analysed using various software choices depending on the instrument used.

In silico peptide mining

Peptides of diverse origins serve various roles in nature, and therefore may bear various modifications. The modifications pose a particular difficulty for peptide analytics and characterization, and the taxonomic origin (Fig. 1) requires adapted/tailored research strategies. In the post-genomic era, scientists have access to databases[77–82] and tools to address these issues (Table 1). Most publicly available ab initio gene and protein sequence data are annotated by programs such as GeneMark[83] or Prodigal[84]. These are robust platforms but come with their limitations as they do not always annotate short open reading frames (sORFs) where peptides are often found. For that purpose, specialized tools exist with their own rule sets for in silico peptide mining. They can be specialized to predict multiple post-translational enzymes, amino acid substrates or biosynthetic tailoring of non-ribosomal peptides, assisting the researcher in annotating possible peptide modifications. In the case of eukaryotic organisms, bioactive peptides may originate from sORFs or derive from breakdown products from other enzymes[58,85,86]. For eukaryotic genetically encoded peptides, there are tools such as SPADA[87], MiPepid[88], DeepCPP[89] or rAMPage[90] (Table 1 and Supplementary Table 1) if genomic or transcriptomic data are available. For protein-derived bioactive peptides, PeptideLocator[86] can be used on protein sequences. Alternatively, comparative genomics can be helpful with tools such as CoGe[91] or EDGAR[92] (Table 1), which may be used when searching conserved homologues for validation; for a summary of databases and software tools, refer to Supplementary Table 1.

The biosynthetic genes of ribosomally synthesized and post-translationally modified peptides (RiPPs) and non-ribosomally synthesized peptides (NRPs) from bacteria and fungi[93–95] are commonly encoded in biosynthetic gene clusters and require programs specialized in biosynthetic gene cluster detection, such as antiSMASH[95] and DeepBGC[96] (Tables 1 and 2). These tools can be complemented by phylogenetic genome mining using EvoMining[97] for the discovery of homologous gene clusters. There are even more specialized tools for NRPs, such as SANDPUMA[98], for the prediction of the substrates of the adenylation domains (A-domains). Furthermore, there are other mining tools for RIPP analysis, such as BAGEL[99] or RODEO[100] for detection and classification of RiPPs from the genome or DeepRiPP[101] for classification, structure prediction and spectral assignment (Tables 1 and 2). These programs use genomic or transcriptomic data to discover

Table 1 | Selected tools for peptide mining

Tool	Organism	Data type	Description
Sequence annotation			
SPADA	Plant	DNA	sORF annotation
MiPepid	Eukaryotic	DNA	sORF annotation
rAMPage	Eukaryotic	RNA	Antimicrobial peptide detection
DeepCPP	Eukaryotic	RNA	sORF annotation
antiSmash	Bacterial/fungal/plant	DNA	Gene cluster annotation
RODEO	Bacterial	Amino acid	Gene neighbourhood analysis
BAGEL	Bacterial	DNA	Gene cluster annotation
DeepBGC	Bacterial	DNA	Gene cluster annotation
Comparative genomics			
EvoMining	Bacterial/fungal/plant	Amino acid	Phylogenetic gene cluster search
CoGE	Eukaryotic/prokaryotic	DNA	Genome comparison
EDGAR	Eukaryotic/prokaryotic	DNA	Genome comparison
Product prediction			
SANDPUMA	Bacterial	Amino acid	A-domain specificity prediction
Peptidelocator	Eukaryotic	Amino acid	Prediction of bioactive peptides derived from enzyme degradation
Multi-omics			
Nerpa	Bacterial/fungal	DNA/mass spectrometry	Maps NRPs back to their respective gene clusters
BioCAT	Bacterial/fungal	DNA/mass spectrometry	Maps NRPs back to their respective gene clusters
DeepRiPP	Bacterial/fungal	DNA/mass spectrometry	Structure elucidation of RiPPs from mass spectra and sequence data
PoGo	Eukaryotic	DNA/mass spectrometry	Peptidogenetic tool, mapping peptides to the genomic loci
MetaMiner	Bacterial	DNA/mass spectrometry	Large-scale screening platform for microbial peptides

In silico peptide mining is aided by a myriad of interesting tools from sequence annotation, comparative genomics, product predictions and multi-omics approaches. However, a tool needs to be matched with its proper use, as the genes and genomic architecture of plants, animals, bacteria or fungi are not the same, and will therefore often need detection rules, dedicated to each clade of life. The detection and discovery of potential peptides can be done at each level of sequence data from genomic (DNA) to transcriptomic (RNA) or by further investigation into proteins (amino acid) for their degradation products, or neighbouring genes may be bioactive peptides or involved in bioactive peptide processing. Furthermore, comparison of sequence data with mass spectra has led to the development of robust multi-omics platforms to aid researchers in high-throughput peptidomics. A-domain, adenylation domain; NRP, non-ribosomally synthesized peptide; RiPP, ribosomally synthesized and post-translationally modified peptide; sORF, short open reading frame.

potential peptides. Alternatively, tools such as PoGo[102] map detected ribosomally synthesized peptides from LC-MS data to the genome. Recently, a retro-biosynthetic tool for NRPs has been developed[103], which allowed non-ribosomal peptides to be mapped back to their respective gene clusters by tools such as Nerpa[104] and BioCAT[105]. The process of discovering novel peptides is an iterative process of in silico and laboratory work, where new discoveries constantly feed the expanding databases, allowing for more precise and detailed tools to be developed (Fig. 2).

Results

The comprehensive sequence identification of peptides includes the full assignment of amino acids in the correct sequence orientation, which is usually determined by the encoding genes and/or the ribosomal protein translational machinery of a cell. The identification and the site of post-translational modifications or tailoring reactions are an additional challenge to allow full assignment of the native peptide. This section provides an overview on the workflows and frameworks that have been implemented to analyse different sets of peptide analytes using peptide mapping and de novo sequencing, automated LC-MS/MS workflows and MSI (Fig. 3).

Peptide mapping and de novo sequencing

Despite the use of coupled LC-MS methods, for certain applications, such as complex peptide analytes, an offline workflow can be beneficial. For such applications, MALDI-TOF/TOF-MS[41] is often the method of choice as it results in spectra with singly charged ions, for example $[M + H]^+$, suitable for manual peptide annotation and de novo sequencing (Fig. 3a).

Peptide dereplication and peptide mapping. Peptide dereplication is commonly applied for rapid pre-screening of peptide libraries, for example peptide natural products in extracted samples from microbial or plant origin[106]. Dereplication can be achieved at different levels, for example by matching experimental *m/z* signals (mass spectrometry) or spectral MS/MS data to libraries. Peptide *m/z* signals provide valuable information for peptide content mapping by comparison of those experimentally determined with calculated molecular masses in databases. HRMS data further enable comparison of isotopologue intensities with theoretical data-based intensities as well as prediction of chemical sum formula, which can provide a further layer of evidence for the matched library hit. Despite the lack of comprehensive databases targeting natural product/microbial-derived

peptides, they are listed in common natural product databases, such as the Dictionary of Natural Products (Taylor & Francis) or Antibase[107] (others are under development), or specialized databases such as Bactibase dedicated to subsets or subclasses of peptides (Table 2 and Supplementary Table 1). The search runs are performed by using computational-aided tools for assignment and further evaluation. DEREPLICATOR+[108] (Table 2), SEQUEST[109] and MZmine3 have emerged as helpful tools. For example, the MZmine3 engine offers an all-in-one solution for peak picking, chromatographic peak detection, peak identification and quantitation, and data processing/visualization functionalities[110]. Using high-resolution mass spectrometers, such as FTICR-MS or orbitrap devices, it has become possible to derive molecular formulas of peptides based on their determined accurate mass. Common tools for molecular formula determination based on high-resolution spectra are pacMass[111], MS-FINDER[112] (Table 2) or MetaSpace[113]. Moreover, spectral library search approaches are also advancing into the peptidomics field as, besides customized in-house spectral databases, commercial (for example, mzCloud) or open access (such as NIST) spectral libraries are readily available for mapping experiments. A combination of database dereplication, high-resolution mass detection and spectral database annotation is state of the art for peptide mapping approaches today. Several manufacturers of mass spectrometers offer software solutions, such as Metaboscape (Bruker Daltonics) or Compound Discoverer (Thermo Fisher Scientific), to perform the identification of peptides using one or several of the described methods in one package. The annotation of branched or cyclic peptides was addressed by CycloBranch2 (refs. [114,115]) or the VarQuest algorithm using spectral networks[116]. As a general limitation, a comprehensive peptide database including high-resolution accurate mass, sequence information, fragmentation spectra is still not available to the research field to date.

De novo sequencing. MALDI-TOF/TOF systems are capable of post-source decay fragmentation. Owing to metastable decay, the system offers less efficient fragmentation than collision-induced dissociation or electron transfer dissociation methods. This fragmentation technology provides signal intensities for specific fragment ions, which are useful for manual de novo sequencing applications[117] (Fig. 3a). De novo sequencing examines characteristic mass shifts among the fragment ions to reconstruct ion series indicating non-random amino acid combinations, thus allowing for the detection of novel sequences[118]. De novo interpretation of fragmentation spectra is still the best choice to derive sequence information of highly functionalized peptides (for example, peptides with post-translational modifications and/or non-natural amino acids). The post-source decay fragmentation approach is useful for studying single to few peptides without much

Table 2 | List of important databases and tools

Databases and software	Use case	Data available/input data	Key features
Database			
NCBI	Biological sequences	DNA, RNA, amino acid	Repository of various biological sequences
Dictionary of Natural Products[a]	Physicochemical data	Solubility, UV–Vis	Extensive database resource of natural products and their physiochemical properties
Software for in silico annotation			
antiSMASH	RiPPs and NRPs	DNA	Rule-based cluster detection
SANDPUMA	NRPs	DNA	A-domain specificity prediction
DeepBGC	RiPPs and NRPs	DNA	Gene cluster detection
Software for peptidogenetic pipelines			
DeepRiPP	RiPPs	DNA (open reading frame)	Classification, processing and spectral matching
Software for mass spectrometry analysis			
DEREPLICATOR+	RiPPs and NRPs	LC-MS/MS data	Natural product identification from mass spectrometry spectra (GNPS framework)
MS-FINDER	Mass spectrometry data analysis	EI-MS, GC-MS, MS/MS	Formula predictions, fragment annotations and structure elucidation
Software for MSI[a]			
MSiReader	Mass spectrometry data analysis	MSI	MSI platform for analysis
SCiLS Lab[a]	Mass spectrometry data analysis	MSI	MSI platform
ImageQuest[a]	Mass spectrometry data analysis	MSI	MSI platform
High Definition Imaging[a]	Mass spectrometry data analysis	MSI	MSI platform
msiQuant	Mass spectrometry data analysis	MSI	MSI platform for analysis

Various resources can aid in peptidomics and one of the main resources most researchers start with are biological sequence databases, to gather genomic (DNA), transcriptomic (RNA), proteomic (amino acid) or other data that may be relevant to their research. These data types can be further complemented by specialized software or other physiochemical or spectral data to make more accurate predictions or annotations of peptides that may be present in the sample. Databases and further tools are continued in Supplementary Table 1. A-domain, adenylation domain; EI-MS, electron ionization–mass spectrometry; GC-MS, gas chromatography–mass spectrometry; LC-MS, liquid chromatography–mass spectrometry; MSI, mass spectrometry imaging; MS-MS, tandem mass spectrometry; NRP, non-ribosomally synthesized peptide; RiPP, ribosomally synthesized and post-translationally modified peptide. [a]Commercial platforms, may be subject to licensing charges.

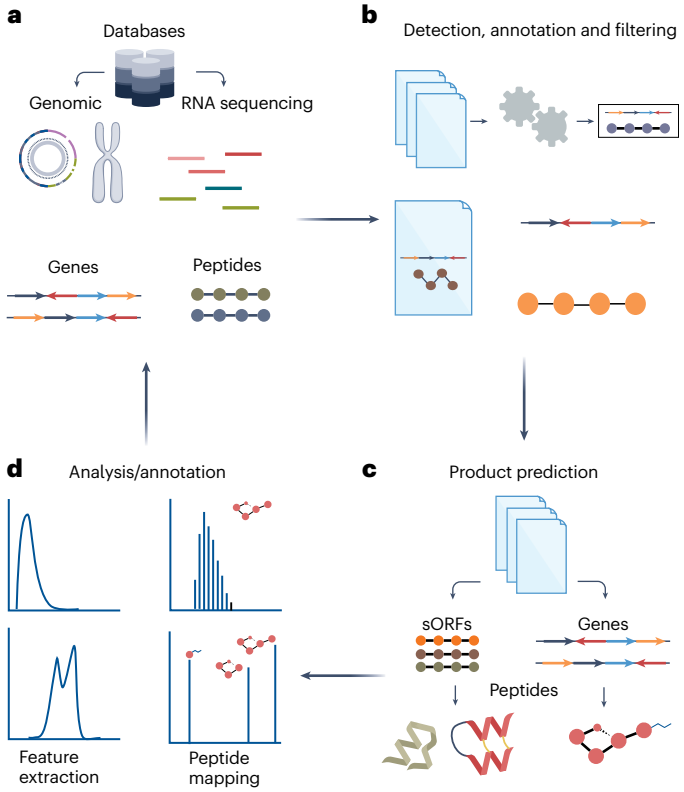


Fig. 2 | In silico peptide mining. General workflow of how multi-omics approaches may be applied to peptidomics. **a**, Novel peptide candidates are often discovered in publicly available data sets, from genomics, transcriptomics or proteins. **b**, FASTA files are fed into the annotation platforms, where they utilize their own detection rule sets to predict the likelihood of true peptide candidates. **c**, Peptides derived from ribosomal transcription may undergo extensive modification, or non-ribosomal proteins may be responsible for peptide biosynthesis; as such, product prediction tools have been made to aid in novel peptide discovery. **d**, Finally, candidates validated by chemical analysis will be deposited in databases, increasing our collective knowledge and data amount to further improve algorithms and bioinformatics tools. sORF, short open reading frame.

need for further automation. High-throughput analysis and combinations together with other omics techniques made a software-assisted solution for de novo interpretation of MS/MS spectra indispensable to handle these large data sets. Bioinformatic tools, such as PEAKS[119], are algorithms for de novo peptide/protein sequencing and allow combinatorial database annotation. An important consideration of de novo sequencing is handling of false positive assignments. Deep learning methods have significantly improved the power of de novo sequencing methods, allowing ≥97% sequence coverage[120]. Although the expected sequence coverages with and without the assistance of databases improved over the past two decades, the algorithms to assign amino acid sequences de novo to experimental MS/MS data are still a source of variation. As these algorithms can substantially differ from each other, for example in their cut-offs for profitability probing or in other decision-making processes, their output for one spectrum can be conflicting with obvious false assignments. To account for this remaining challenge of software-assisted analysis, de novo sequence work should be carefully validated using known or reference peptides whenever possible. The quality of de novo data may also depend on instrument performance, spectra quality, peptide fragmentation efficiency, presence of post-translational modifications, the abundance of precursor ions or sample size[69,121,122].

Peptide databases and identification workflows

Regardless of the acquisition method (DDA or DIA), the output of a typical LC-MS/MS experiment comprises hundreds of thousands of peptide fragmentation spectra correlated to chromatographic retention time and precursor ion mass (Fig. 3b). Interpretation of such spectral data typically relies on querying them against in silico predicted theoretical spectra from protein sequences found in a species proteome database (Table 2 and Supplementary Table 1). More effective, however, is a de novo sequence tag approach that infers the peptide sequence directly from characteristic mass shifts between peptide fragment ions, and then matches the tag to proteins in the database. The advantage of de novo sequencing is its ability to discover peptides outside the proteome database, as unassigned tags can be searched against expressed sequence tag (EST) repositories or mass spectrometry databases of homologous species from other studies. With either matching algorithm, experimental data fit to matched proteins are statistically evaluated for probability and the false discovery rate (FDR).

Peptide identification from DDA spectra is typically done by ranking the probability of non-random fit of peak patterns in the MS/MS spectrum into certain amino acid sequences using mass spectrometry vendor-specific analysis tools or the universal format software PEAKS[7]. Annotation of post-translational modifications is typically included with either option. It is important to recognize that popular proteomics tools MaxQuant[123] and Mascot[124] are not suitable for native peptide identifications as they rely on in silico spectral libraries of theoretical peptides potentially originating from enzymatic cleavages of proteins in a database; when the enzyme is not specified, dramatic expansion of search space overwhelms computational resources.

Owing to a conceptually different fragmentation approach in DIA experiments, alternative software is needed for identification and quantitation of peptides, for example Skyline[125], DIA-NN[50], OpenSWATH[126], Spectronaut[127] and DIA-Umpire[51]. Statistical analysis of DIA measured peptides can be performed with the output result files using Excel, R programming, Python and Perseus[123] (Table 2 and Supplementary Table 1).

Quantitative peptidomics

The application of mass spectrometry-based peptide quantitation is rapidly growing in clinical, applied and basic research. Traditionally, in bottom-up proteomics, the mass spectrometry quantitation approach is based on comparison of protein levels via summation of measurements from several encoded tryptic peptides. In peptidomics, however, individual bioactive peptides may have independent levels in relation to pathological or experimental conditions, even if originating from the same protein or prohormone. Therefore, endogenous peptides that are subject to peptidomics investigation will be quantified individually at the peptide level, not the protein level. With that difference in mind, practical strategies for quantitative peptidomics are similar to those widely used in bottom-up proteomics and include stable isotope or chemical labelling and label-free methods[28–32] (Fig. 4). An advantage of the label-free quantitation approach is its low cost and simplicity of sample preparation. The two commonly used label-free quantitation techniques are based on the signal intensity (using extracted

ion chromatograms) and spectral counting[128]. Both methods can be used for relative and absolute quantitation. For absolute quantitation, a peptide standard that is similar to the target peptide can be added[43], or ideally a synthetic stable isotope-labelled internal standard for each peptide of interest is used, for example AQUA[46] peptide. In silico algorithm-based methods may also be used to achieve absolute quantitation and assess the actual concentration of the target peptide in one sample[58,129,130]. Additionally, multiple reaction monitoring can be used in targeted peptide quantitation. Multiple reaction monitoring focuses on selected fragment ions for peptides of interest and allows the detection of low-abundance peptides[44].

The advantage of the stable isotope labelling strategy is multiplexing and throughput, although often at high cost. The quantitation of labelled peptides is based on the mass spectrometry signal intensity, and the complexity of analysis increases with the size of data sets. Conceptually different in vitro strategies involve isobaric labels[131] that produce identical mass shifts on the mass spectrometry level, but generate distinct reporter ions associated with different labelling channels during peptide fragmentation in MS/MS events[131]. We refer the readers to another article for more extensive discussion on isobaric labelling[131]. Raw files obtained from both label-free and labelling strategies can be analysed through the software mentioned in the previous section and detailed in Table 2 and Supplementary Table 1.

Mass spectrometry imaging and spatial patterning

MSI is a complementary method to IHC staining[60], radio-immunoassays[61] and fluorescence microscopy[62], to yield images of the target analyte in relatively high throughput[53]. Figure 3c demonstrates the experimental procedure and data processing steps for MSI experiments. Distribution of a list of the target *m/z* can be visualized in one experiment from the same tissue slice. Although MSI provides direct qualitative results for the target *m/z*, quantitative analysis can also be done with the aid of commercial and open access software. To relatively quantify the desired *m/z* for a certain peptide, direct comparison can be done among the different tissue regions or different tissue slices[63]. Normalization is usually completed pre and post processing[132]. The software is equipped with normalization tools, using the total ion chromatogram (TIC), vector norm through root mean square, median and noise[133]. Similar to LC-MS, isotopic labelling may be transformed for MSI relative quantitation[134]. Absolute quantitation is a relatively untapped area, and such quantitation can be performed by LC-MS/MS or by adding the calibrant to the solvent stream in desorption ESI MSI experiments[135,136]. Another way to perform absolute quantitation is to create a calibration curve by spotting the standards of interest onto a different tissue section that is adjacent to the analysed tissue sections[137].

Various software tools (Table 2 and Supplementary Table 1) have been developed for MSI visualization and analysis, including MSiReader[138], SCiLS Lab, ImageQuest, High Definition Imaging, MassImager[139] and msiQuant[140]. Several statistical analyses can be performed using some of these software packages, including analysis of variance, principal component analysis and partial least squares coupled to discriminant analysis (PCA-DA and PLS-DA, respectively), and receiver operator characteristic curve analysis for biomarker tests[63]. Subsequent statistical and classification analysis can also be done using machine learning and in silico algorithm-based software. MALDI MSI data are usually paired with LC-MS/MS data for specific peak assignment and peptide identification. We refer the readers to a more thorough review article[63] for a discussion on MSI.

Applications

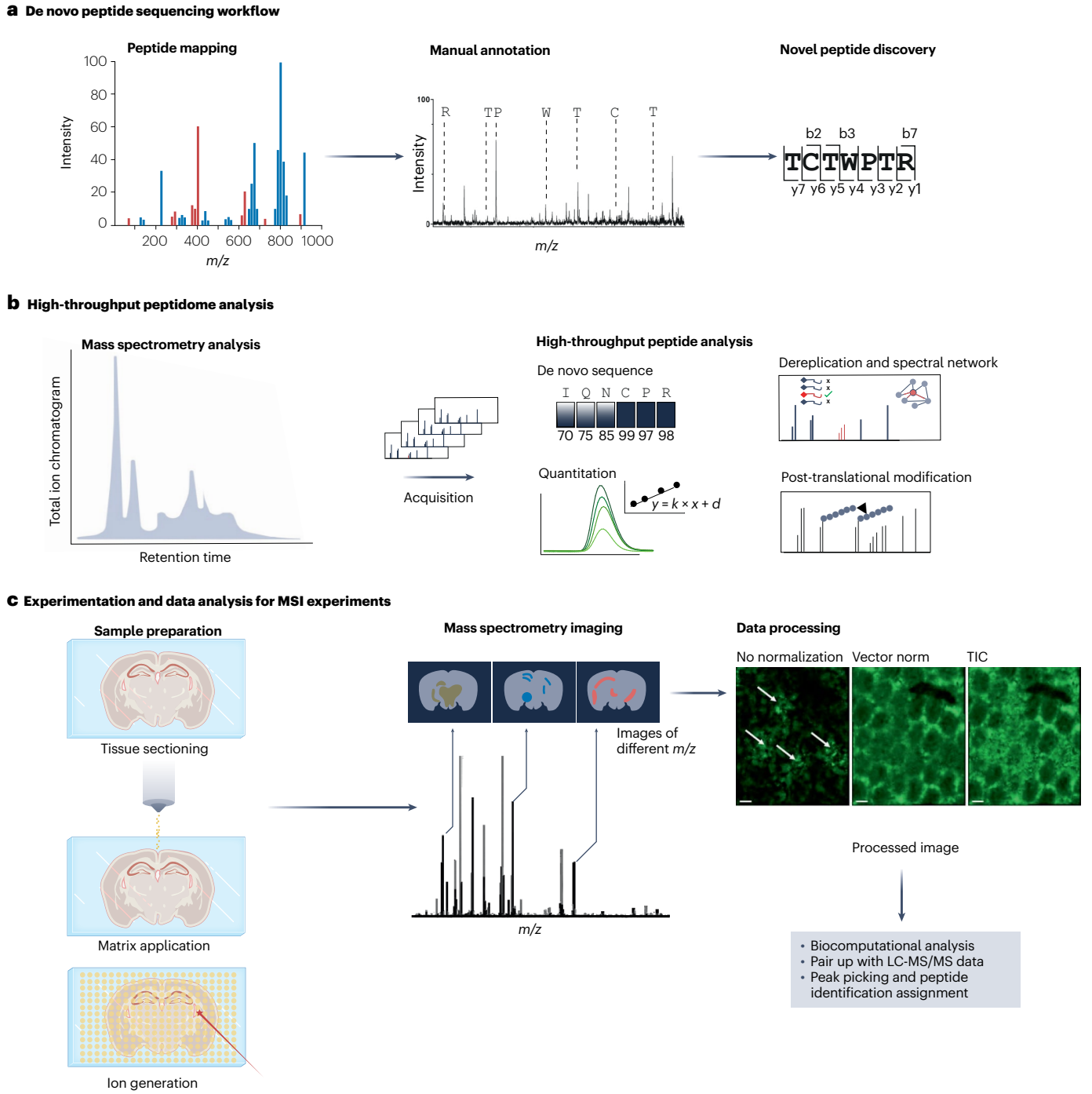
Function, diversity and evolution in biology

Microbial peptide biosynthesis. Some peptide species from bacteria and fungi are distinct from peptides commonly found in higher organisms. These peptides, according to their more specialized biosynthetic mechanisms, can be categorized into two groups: RiPPs and NRPs. RiPPs are encoded as precursor peptides in the genome and undergo post-translational modifications beyond those commonly known from animal/plant-derived peptides and proteins[141]. RiPPs are subcategorized by their characteristic post-translational modifications, which include thioethers, heterocycles, amino acid side-chain functionalizations and various side-chain cross links. Recently, examples such as the lipolanthines emerged where RiPP biosynthesis is combined with other biosynthesis pathways, such as fatty acid biosynthesis of polyketide synthesis (PKS)[142].

By contrast, NRPs are commonly synthesized by modular multidomain synthetases that can incorporate non-proteinogenic amino acids and other substrates. Modules for peptide assembly consist of a basic domain set: A-domains responsible for amino acid/substrate activation and selectivity; a peptidyl-carrier protein domain; and a condensation domain (C-domain) forming the amide bonds. A thioesterase domain (TE-domain) releases a linear or cyclic peptide. Additional domains may be interspersed to perform epimerizations (E-domains), oxidations/reductions (Ox/Red-domains), cyclizations (Cy-domains) or other modifications[143]. Tailoring of the peptide substrate may occur to an extent that, sometimes, a peptide structure is hardly recognized in the final product, for example in the β-lactams (precursor peptide L-aminoadipoyl-L-Cys-D-Val). The structural diversity of NRPs is even further extended by mixing biosynthetic functions with PKS, rendering lipopeptides or even stronger morphed PKS-NRP-like structures, which probably mark the borders of the peptidomics field[144]. Although, recently, genome mining considerably facilitated the discovery of new classes and types of microbial peptides derived from RiPP and NRP biosynthesis, the field still is in its infancy, as estimates consider that only 2.1% of the global prokaryotic taxa are represented in sequenced genomes[145].

Venomics. Venomous animals (including snakes, spiders, scorpions, amphibians, snails and even platypuses) have evolved multiple times throughout evolution, and many venom cocktails are rich in peptides[146–148]. In fact, the diversity of venom peptides is unprecedented: their estimated number exceeds millions[149]. Peptidomics has provided detailed insights into the diversification of venom. The venom content has also evolved, based on the targeted prey or predators to defend against. In research on the predation of cone snails *Conus marmoreus* and *Conus geographus*, their conotoxins and the defence stings showed remarkably different venoms. Whereas defensive venoms are localized in the proximal duct, the predation venoms are in the distal duct of the venom gland[150,151]. Predatory venoms evolved to incapacitate or kill the prey with high selectivity, whereas the defensive venoms had little to no activity against their prey but contained high amounts of paralytic peptides (conotoxins) acting on mammalian ion channels. Scorpions have also been studied extensively for this purpose. Their venom components differ concerning defensive and predatory behaviours; they might sting multiple times, but the venom composition may vary in response to the threat of the animal[152,153]. These examples underline the requirement for appropriate extraction methods when working with venomous animals.

Invertebrate neuropeptide discovery. There is an overwhelming diversity of (neuro)peptidomics studies on invertebrates, including



the classic model species in insects[154–156], molluscs[157–160], worms, crustaceans[161–165] and cnidarians[166–168]. Multiple approaches proved effective for discovery and characterization of neuropeptides in invertebrates. With growing numbers of sequenced genomes or transcriptomes, the bioinformatics annotation of prohormones and the prediction of endogenous putative peptides facilitates peptidomics studies[169]. The characterization of structurally simple peptides can be achieved by matching the mass spectrometry-detected peptide masses to theoretical masses of peptides predicted from a protein/prohormone sequence while accounting for possible post-translational modifications[170]. This is an effective approach suitable for single-cell peptidomics[171], and thus allowing for the discovery of chemical messengers in well-defined neuronal circuits[172] that control physiological functions and behaviour, for which invertebrates are extremely well suited due to the simplicity of their organization and conservation of signalling molecules along the evolution tree. For characterization of the peptidome of an organism with no genomic information, shotgun peptidomics on a larger tissue sample works best with mass spectrometry data annotated using the

Fig. 3 | Overview of common peptidomics workflows. a, De novo peptide sequencing is utilized in peptide discovery research. Purified peptides are analysed with matrix-assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF-MS) or tandem mass spectrometry (MS/MS) systems. **b**, High-throughput peptidome analysis is often employed for automated sequence analysis, dereplication and/or spectral network analysis, quantitation and analysis for post-translational modifications. These analyses are usually paired with high-throughput analytical systems such as high-resolution liquid chromatography–tandem mass spectrometry (LC-MS/MS). Data acquisition generates MS/MS spectra that can be analysed with de novo sequencing algorithms or used for identification of peptides with a database search. Quantitation of peptides can be achieved at the level of the mass spectrometry peak as well as with multiple reaction monitoring acquisition. Post-translational modification analysis can be performed for tryptic peptides or for endogenous peptides using a peptide fragment spectrum for database search. **c**, Experimental protocol and data analysis procedure for mass spectrometry imaging (MSI). After sample preparation of tissue slices (for example, mouse brain tissue), the matrix is applied and co-registered with a tissue image to precisely define sample position in the mass spectrometer. The data acquisition over the entire raster generates a summed total ion chromatogram (TIC) spectrum. To visualize *m/z* maps, each *m/z* signal can be selected to show extracted ion chromatograms. The data can be processed (normalization, smoothing, data compression) for various biocomputational analysis. Scale bars, 200 µm. Part **c** data processing images reprinted from ref. [133], Springer Nature Limited.

protein database of a related species (homology search)[173]. Specifically, hybrid bioinformatics approaches for interpretation of the MS/MS data match experimental de novo sequence tags to a database of related species proteins while accounting for potential single point mutations[174]. A more targeted analysis aimed at peptide-level validation of selected gene expression and/or function in specific cells or tissue often involves a multi-omics workflow[175] or gene cloning followed by gene expression mapping to guide tissue or cell sampling for mass spectrometry analysis[176–179]. Elegant multi-platform studies combining shotgun peptidomics for peptide library construction followed by MSI of tissue sections have led to mapping of putative bioactive peptides in nervous system tissue sections under different biological paradigms, as well as the exploration of cellular heterogeneity and organelle peptide complements. Implementation of mass spectrometry-based, peptide-centred workflows has accelerated the discovery of enzyme-derived D-amino acid-containing peptides (DAACPs) of high physiological importance in invertebrates[176,180,181]. DAACPs typically co-exist with all-L-amino acid-containing peptide counterparts in tissue extracts and can be fractionated via RP-HPLC and can be validated via trapped IMS and MS/MS[182,183]. An alternative discovery pipeline based on enzymatic screening, separation and amino acid analysis is greatly enhanced by MS/MS[184–186].

Human and mammalian neuropeptide discovery. Peptidergic systems are abundant ligand–receptor signalling systems in mammals and are of high interest to further understand human signalling networks. There are numerous peptide hormones in the body but their receptor targets, many of them from the G-protein-coupled receptor family, and the physiological role of these systems remain elusive. Peptidomics assisted the deorphanization process of peptide/protein target signalling networks in the past. Many neuropeptides and their endogenous receptors were discovered with bioinformatics tools, as the human genome sequence data enabled large-scale analysis. Conceptionally, bioinformatics has limitations to detect precursor splicing, secretion signal sequences and post-translational modifications. Peptidomics approaches were applied in neuropeptide discovery but have limited sensitivity for the detection, and the identification is often knowledge-based using pre-defined search libraries[187]. Recent developments highlight a robust analytic framework for extracting, analysing and identifying endogenous peptides[188]. Integrated computational and experimental approaches have been powerful tools for peptide–orphan GPCR pairing[189,190]. Human peptide ligands of orphan receptors were predicted based on common sequence motifs, for example secretion sequences and conserved regions in their encoding precursors. The identified peptides can be chemically synthesized for testing in vitro for activation of receptor systems. These deciphered systems revealed new peptide ligands for the GPR1, GPR15, GPR55, GPR68 and BB_3 receptors[189]. The human neuropeptide discovery field is hugely significant to understand physiological process and pathophysiological conditions, providing clinical opportunities for new therapies of brain disorders. For example, combinatorial workflows to enable large-scale mass spectrometry-based peptidomics for drug discovery or integrated workflows to decipher signalling systems provided significant contributions[188,189]. Overall, there is a need for in-depth studies of the human neuropeptidome, which is still at the frontiers stage compared with other omics technologies.

Drug discovery

From venoms to drugs. Historically, the discovery of drugs has been accomplished by natural observation, followed by trial and error experimentation with various plants and animal extracts[191,192]. Traditional knowledge, passed on by generations, became the pillar of drug discovery as medical sciences were established and methods were developed to validate the bioactivity[193]. Modern drug discovery can be split into two phases: compound screening at the molecular targets (Fig. 5a,b) and medicinal chemistry efforts (Fig. 5b,c) to improve the pharmacological properties[194]. Peptidomics has played a crucial role in the discovery of peptide drugs and peptide-derived drugs. A prominent example is the bradykinin potentiating peptide isolated from the Brazilian pit viper *Bothrops jararaca*; the initial peptide isolated was the template for developing a small-molecule peptidomimetic resulting in the angiotensin converting enzyme inhibitor captopril, a hypertension medication[195] (Fig. 5c).

Owing to their neurotoxic effects, peptide analgesics are known to be common components of venoms. Peptidomics analysis has shed light on the composition and structures of these highly complex peptide mixtures. Therefore, it has been possible to isolate and identify peptides from various animals such as scorpions[196], cnidarians[197] and cone snails[198]. One such example, the cone snail conotoxin MVIIA is an N-type channel blocker. The synthetic version, marketed as Ziconotide (Prialt), is used to treat chronic pain[199]. Venom peptides also target peptide hormone systems. Remarkably, cone snails have weaponized peptide hormones by generating fish insulin analogues, releasing them from their venom glands and sending their prey into hypoglycaemic shock[200]. Peptidomics was the key technology to isolate and identify the activity-bearing peptides. The last example worth mentioning is the Gila monster (lizard) peptide exendin 4. It is a long-acting GLP1 mimetic, which led to the development of the GLP1 antagonist exenatide[201] (Fig. 5c) preceding liraglutide and semaglutide as peptide drugs in the same therapeutic area[202]. GLP1 and GIP are endogenous peptide

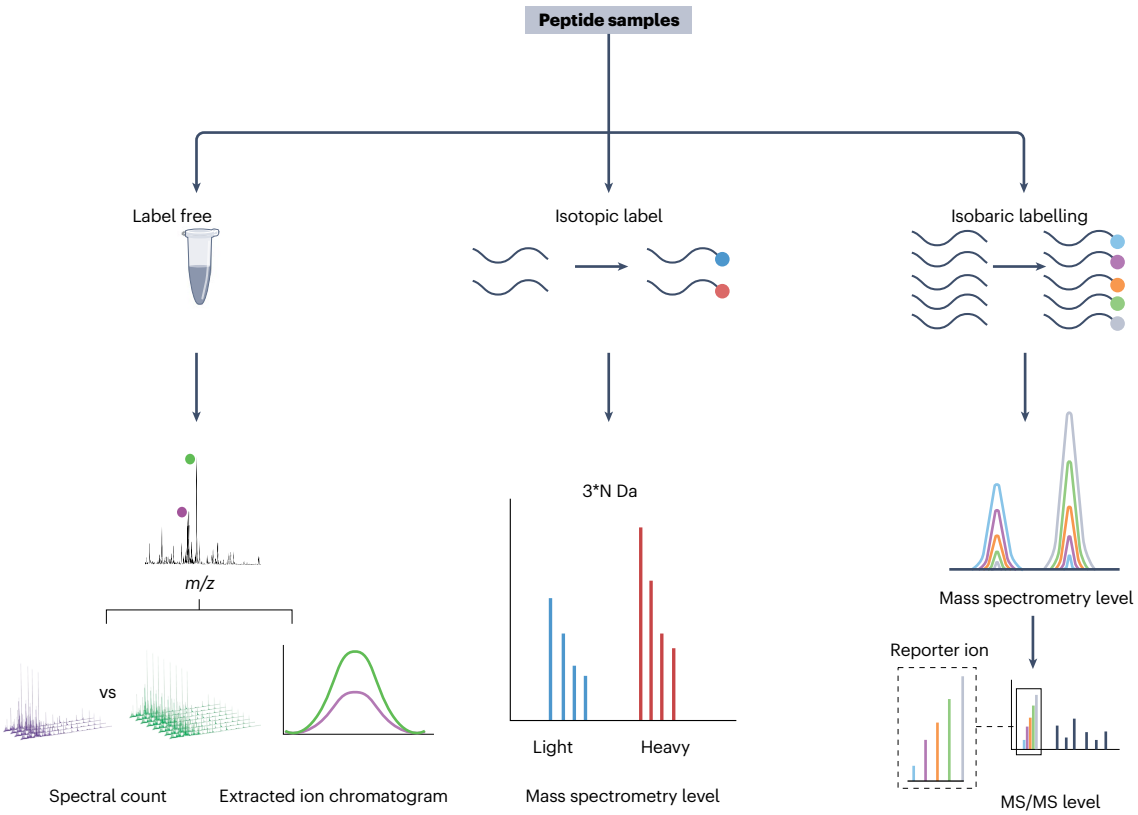


Fig. 4 | Quantitation methods for peptidomics. Techniques include label-free, isotopic labelling and isobaric labelling quantitation. Label-free quantitation (left panel) is based on the mass spectrometry level and can be achieved through either spectral count or extracted ion chromatograms; labelled strategies can be achieved at both the mass spectrometry and tandem mass spectrometry (MS/MS) levels. Samples labelled with an isotopic labelling strategy (middle panel) can be quantified through the mass defect on the mass spectrometry level, whereas isobaric-labelled samples (right panel) are quantified on the MS/MS level. Bottom panel reprinted with permission from ref. [165], Wiley.

hormones, which amplify the response of glucose-induced insulin secretion. Hence, these venom-derived and hormone-derived peptides, identified with the help of peptidomics, are being used clinically as antidiabetic drugs.

Antimicrobial peptides and derivatives. Although possibly not regarded as antimicrobial peptides in the classical sense, microbial peptides and derivatives have been used for decades as highly successful, mostly anti-infective drugs. Examples of marketed NRP-derived peptide drugs, which all show sophisticated mechanisms of action, are vancomycin[203], daptomycin[204], bleomycin[205] (anticancer), cyclosporin[206] (immunosuppressant) and even β-lactams. Among RiPPs, nisin (E234) is the frequently mentioned example serving as a food preservative[207]. Antimicrobial peptides in a narrower sense include defensins, which kickstarted the discovery of peptides belonging to the innate immune system[208]. Antimicrobial peptides originate from multicellular organisms (fungi, animals and plants) and range between 10 and 120 amino acids in size. The assignment to this family is very broad, preferably with an overall positive charge and cross-linked by disulfide bridges[209], but other structural types are known. Antimicrobial peptides often affect cells by rupturing the lipid layer, pore formation or inhibiting cell-wall synthesis[205]. Recent peptide discoveries with impressive antibacterial activity against relevant Gram-negative bacteria include albicidin (NRP)[210] and darobactin (RiPP)[211], which may contribute to the much-needed demand for new antibiotics.

Body fluid peptidomics — novel antiviral drug candidates. Although it is established that peptides of the innate immune system possess antibacterial and antifungal properties, these peptides also possess antiviral or anticancer properties[6,212,213]. These functions are frequent in biology and are termed moonlighting activities. The features that improve antiviral properties correlate to the cationic charge and the peptide's hydrophobicity. The peptides directly bind to the viral particles themselves, preventing viral fusion with the host cell. These peptides display characteristics to prevent or to encounter viral infections[6,214] and further development may eventually lead to future drugs.

Clinical applications

Biomarker discovery. Mass spectrometry has made significant contributions to identify and validate potential peptide biomarkers. By comparing healthy and diseased tissues or body fluids, differential display of the endogenous peptides can indicate potential biomarkers or a therapeutic target for a disease. The development of improved mass spectrometry techniques enabled the discovery of low-abundance peptides in clinical samples, especially for peptides from biofluids. Many studies utilized sensitive mass spectrometry tools to investigate

the urinary peptidome for kidney-related diseases[215]. The urinary peptidome was investigated with capillary electrophoresis–TOF-MS and further verified with capillary electrophoresis–FTICR-MS, where 273 peptides were identified to be associated with advanced chronic kidney disease[216], which later proved to be biomarkers for chronic kidney disease progression and diabetic nephropathy[217,218]. Instead of the traditionally used positron emission tomography technique[219], researchers have also leveraged MALDI MSI to study diseases with analyte localization information. Exemplarily, the spatial progression of amyloid aggregates for Alzheimer disease was investigated through multimodal MALDI MSI[220].

Targeted characterization and quantitation of peptides with specific post-translational modifications can also be highly valuable. Among the diverse modifications, glycosylation has attracted increased interest[221], due to its close association with neurodegenerative disorders[222], cancer[223] and autoimmune diseases[224]. The in-depth characterization and quantitation of glycosylated peptides remain challenging due to their low abundance in vivo and high chemical complexity and structural diversity[221]. Many mass spectrometry-related methodologies for glycosylated peptide detection have been reported, including separation and enrichment of the glycopeptides during sample preparation, enhanced fragmentation techniques

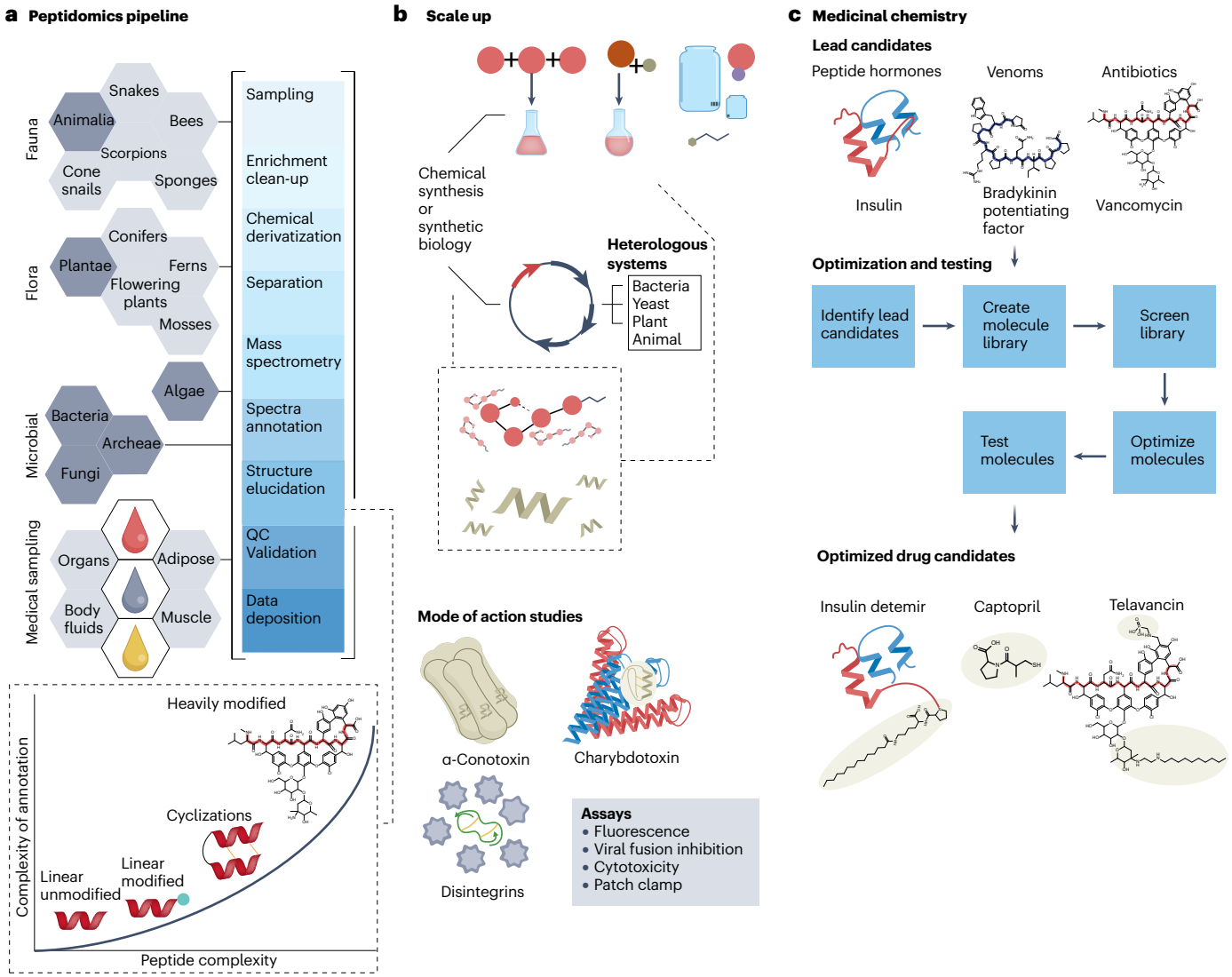


Fig. 5 | From nature to medicine. a, Generalized scheme of peptide sources and the peptidomics analysis pipeline. During structural elucidation, the complexity of spectral annotation will increase with the structural complexity of the peptides (for example, by post-translational modifications such as disulfide bond formation, cyclization). **b**, After peptide analysis, desired products are scaled up by chemical synthesis or synthetic biology (heterologous expression in various host organisms) for mode of action studies and pharmacological analysis. **c**, Optimization of the lead peptides (for example, insulin, bradykinin potentiating factor or vancomycin) by medicinal chemistry. Optimization is a multi-step process, including pharmacophore analysis, chemical synthesis, molecule library preparation and screening, structure–activity studies and pharmacology to generate optimized peptide or peptidomimetic drug candidates (for example, insulin detemir, captopril or telavancin) for applications in medicine. Structural alterations during lead to drug optimization are highlighted by pale yellow ovals (peptide backbones are coloured blue and red).

Glossary

Fourier transform ion cyclotron resonance mass spectrometers
(FTICR-MS). High-resolution mass analysers that trap ions in a cyclotron radius by applying a fixed magnetic field and an oscillating electronic field. As the ions rotate, an interferogram signal is recorded by electrodes and the useful mass spectrum is extracted with a Fourier transformation.

Hyphenated front-end separation platforms
Platforms that separate the analytes online before they enter the mass spectrometers. Techniques include, but are not limited to, liquid chromatography, gas chromatography, ion mobility spectrometry (IMS), solid-phase extraction (SPE) and capillary electrophoresis.

Ion mobility spectrometry
(IMS). An analytical technique that sorts and separates gas-phase ions based on their mobility in a carrier buffer gas under the influence of an electrical field, which is related to the conformation and 3D shapes of the molecules.

Multiple reaction monitoring
A type of analysis for tandem mass spectrometers providing capabilities for quantitation of analytes. Pre-defined precursor ions (m/z) are selected by the first mass analyser and submitted to a fragmentation, and the selected product ion m/z signals are detected by the second mass analyser.

Peptide dereplication
Refers to the identification of known peptides in a sample by comparing mass spectrometric data with a library. The identification can be obtained by comparison of m/z mass signals, including the isotopologue intensities and pattern of isotopologues, giving information on the chemical composition as well as on tandem mass spectrometry (MS/MS) fragmentation spectra match with library data.

Peptide spectrum match
(PSM). A scoring function in which the mass spectrum of a peptide is compared with a theoretical peptide sequence to determine the probability of the measured peptide matching the theoretical peptide.

Post-source decay
A type of fragmentation technique that applies when metastable ions spontaneously decompose in the drift region between the ion source and reflectron.

Short open reading frames
(sORFs). Open reading frames that occur throughout the genome and usually comprise <100 codons. They are a possible source for peptides with biological relevance.

(for example, collision-induced dissociation, electron transfer dissociation, EThCD)[47–49] and DIA-MS[139,225]. As an example, a targeted mass spectrometry approach was employed with oxonium ion-triggered EThCD to achieve the first large-scale discovery of O-glycosylation on signalling peptides in human and mouse pancreatic islets[226].

Pathophysiology/physiology: mechanism of disease and treatment. Since the discovery of insulin, neuropeptides and other peptide hormones have been regarded as an important class of chemical regulators broadly involved in mediating numerous physiological functions[227]. Various animal models relevant to human physiology and pathophysiology provided an opportunity to link neurochemical changes to behavioural output and extrapolate findings to humans. Mass spectrometry played a significant role in the discovery and characterization of signalling peptides in evolutionary conserved pathways governing homeostasis[225], pain[228], complex behaviour[229], learning and memory[230], and ageing[231,232], to name a few. High-throughput LC-MS-powered inquiries of animal peptidomes provided molecular links between native peptide dynamic states and environmental factors, nutrition, disease and behaviour[233]. As an example of a tour de force discovery effort, a label-free LC-MS approach was employed to identify and measure neuropeptide levels in a murine migraine model[234]. From 1,500 neuropeptides screened, 16 were linked to migraine and/or opioid-induced hyperalgesia[234]. To focus on secreted peptides, peptidomic analysis can be performed on synaptoneurosomes, dense core vesicles[235] or captured single-cell releasates[36] to probe neuropeptidome dynamics[236], synaptic dysfunction and brain neurodegeneration. A more elegant but difficult approach is an in vivo measurement of secreted peptides via microdialysis coupled to mass spectrometry platforms for identification[237]. Another way of gaining insights into intercellular communication is selective in vitro analysis of secreted, physiologically relevant endogenous peptides released from neuronal networks in response to physiological stimulation, which can be achieved via microfluidic devices[238]. Microfluidics integration with mass spectrometry provides capabilities for molecular structural characterization and label-free and absolute quantitation of peptides[239].

Reproducibility and data deposition

The field of peptidomics deals with highly variable sources of peptides, requiring various extraction methods, clean-up, derivatization and different approaches for analysis. If automated methods are used to assist the analysis, researchers will need to report the FDR. The FDR is a statistical method for determining the rate at which type 1 errors occur in null hypothesis testing. The FDR provides the global confidence of the data set, in contrast to the P value of a peptide spectrum match (PSM) which refers to the percentage likelihood of incorrect assignment. For the FDR estimation, the decoy database is the null hypothesis. Accordingly, the FDR is the number of hits from the null hypothesis (decoy) divided by the number of total hits, providing a global confidence of the data set. Although the P value only accounts for a single PSM and the FDR for the global data set, methods to exclude certain PSMs are termed controlling procedures. The simplest would be the q value, often interpreted as the minimum posterior probability of the null hypothesis or the FDR, which means the FDR and the α threshold are the same. Then, if set at 1%, all PSMs with $P \geq 0.01$ will be rejected[240]. This method may not be sufficient, as many algorithms try to improve the FDR along different parameters using the posterior error probability, which can depend on the length, charge and modifications[241]. Other approaches to controlling the FDR may include using P values, covariates, z scores or the family-wise error rate (FWER)[242]. Reporting how the FDR is controlled is important to any omics approach, along with the null hypothesis of the experiment. This allows the user to answer questions such as whether the incorrect PSMs are truly incorrect if post-translational modifications prevent correct assignment or whether peptides in the experiment are present in the database file. As such, the availability of annotated peptide sequences is an invaluable resource for researchers, and they are encouraged to deposit them in relevant databases.

For reproducibility and traceability, discovered peptides and their modifications should be deposited in open access databases. The repositories of the National Center for Biotechnology Information (NCBI)[78], the European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI)[243] and the DNA Data Bank of Japan (DDBJ)[244] share access to data between themselves under the FAIR (Findability, Accessibility, Interoperability and Reusability) guidelines[245].

The EMBL-EBI has three portals for submission relevant to peptidomics research: SPIN (Edman degradation or manual interpretation of MS/MS spectra)[246], ENA (nucleotide translations of protein-level data)[82] and PRIDE[247] (for sequences identified using search engines). Raw and unprocessed mass spectrometry data can be deposited along with processed protein and peptide output files. The use of a vendor data file format is less common. Instead, open source mzXML (or similar) is standard. The data collection should also contain meta files, such as sample preparation protocols and device settings, as well as information on the sample origin (cell type, tissue). The data repositories have detailed requirements and provide a unique identifier to connect publications with the deposit data files[247]. Specifically, in the case of microbial RiPPs and NRPs (and their corresponding gene clusters), the resource MiBIG relies on very simple inputs of the GenBank accession for all genes in the gene cluster and the SMILES for the compound discovered[81].

Limitations and optimizations

Peptide degradation

Sampling that truly reflects the in vivo state of the tissue is imperative for finding potential biomarkers and regulators, but this is not entirely possible. Native peptides in freshly isolated biological samples are subject to a multitude of interferences including biased sampling, variable sample stability and fast degradation that make measurement and identification of endogenous peptides more challenging relative to traditional bottom-up proteomics. Enzymatic degradation of ubiquitous proteins is particularly detrimental to mass spectrometry analysis, as degradation products fall into a typical peptide mass range, thus obscuring the detection of native peptides typically present in much lower amounts in tissue extracts or biological fluids. To prevent enzymatic protein degradation during tissue sampling, several tissue stabilization approaches have been implemented[248], with heat stabilization being one of the most effective methods of sample preparation[249]. Heat stabilization arrests the ex vivo peptidase activity, thereby conserving the chemical composition of the sample[250].

Biological variation versus sensitivity

Biological systems involve a complex interplay of the organism and its environment, and much of that context is lost in the laboratory setting. Sample collection and culturing of microorganisms, for example, only covers a tiny fraction of the strains that were present in the original sample as cultivation conditions are not known for many strains. However, peptide extraction and analysis without propagation may reveal novel peptides or keep some biosynthetic gene clusters active[251,252]. The same applies for clinical settings: every individual has their own genetics, lifestyle and risk factors that affect health. For instance, where routine clinical diagnostics did not suffice, a novel multi-omics approach managed to identify *Bacteroides vulgatus* proteases as a novel risk factor for ulcerative colitis by adding metapeptidomic data to their analysis, which has often been ignored in clinical practice[253]. The use of multi-omics approaches in conjunction with peptidomics will help address the biological variability and improve the sensitivity by proper sampling and analysis. This might become important, for instance, in personalized medicine applications.

Analysis, algorithms and big data

Nowadays, there are several analysis tools available (Tables 1 and 2 and Supplementary Table 1). Peptide assignment and annotation from genomic data are straightforward if genome data of the organisms of interest are available. However, in silico-derived and mass spectrometry-based data sets are not necessarily identical in routine analysis[254]. These discrepancies are due to ambiguities, such as technical shortcomings in mass spectrometric detection and data processing or because of false positive peptide assignment and annotation. For instance, for certain organisms, such as bacteria and fungi, it can be challenging, due to the small reading frames and/or massive biosynthetic modifications. Furthermore, in silico workflows are biased towards well-known examples and prone to propagating bias[255]. Continuous research on such peptides aids training of the algorithms to improve peptide identifications in the future[95,96,99,100]. Analysis can become cumbersome if access to genomic or transcriptomic data is limited. As sequencing costs continue to decrease and sequencing power continuously increases, more genome data will become available in the future. Furthermore, other sources of peptides are breakdown products of proteins, for example the opioid peptides and haemorphins derived from haemoglobin[58]. This is being addressed by deep learning algorithms being developed to detect bioactive peptides in protein sequences[86].

A special consideration is required for RiPPs and NRPs from bacterial and fungal sources. Previously, classical screening approaches were based on the taxonomic characterization of the strain and the subsequent workflow of analytical and assay techniques. Nowadays, massive bioinformatic data from genome sequences, however, made genome mining the dominating technique: genes or gene clusters are analysed by predictive tools for their putative function, which are subsequently validated experimentally. Precedence of structural and functional data eases assignment to biosynthetic classes. Particularly, in the RiPP field, the increasing availability of DNA sequences in the databases led to a massive boost in the discovery of putative but also new structures. NRPs, where the amino acid sequence is not encoded in the mRNA sequence, are a special case. To predict the potential product of these synthetases, the A-domain specificities[98,256] of subdomains involved in substrate recruitment are used. However, this procedure may fail for new amino acid motifs or for complex-type synthetases if the co-linearity rule is violated. Thus, sequence data on the synthetases alone often do not suffice to predict the structure of the searched peptide[2]. This can be overcome with an increase in known NRP structures and their biosynthetic gene clusters. Although most commercial platforms can do an excellent job on properly linearized and derivatized peptides, recently algorithms have been developed for more complex peptides of microbial origins to be annotated directly from the spectra[101,257]. Machine learning[258] and spectral networks have been explored to identify and assign the chemical nature of peptide natural products[259].

Outlook

Method and instrumentation developments

Rapidly evolving mass spectrometry instrumentation opens new opportunities for in-depth interrogation of peptidomes even in small-volume samples. The latest two-step methodology integrated an ion mobility with TOF or orbitrap mass analysers, leading to unprecedented sensitivity and the highest quality of peptide sequencing. Implementing ion mobility as an additional dimension of separation resulted in improved peptide identification rates, enhanced peptide coverage and greater confidence of post-translational modification assignments[260]. Even unique post-translational modifications such as isomerization that were difficult to deduce by other mass spectrometry methods due to a lack of characteristic mass shifts now can be unambiguously

measured and validated. This technology may attract more attention in the peptidomics field soon. The ion mobility separation can be further combined with MSI to enable the investigation of spatially resolved peptidomics in a high-throughput manner with enhanced chemical information.

Data deposition, open source, unity

Standardization of data deposition and annotation has improved in the past few years with centralized databases such as the NCBI[78], the EMBL[243] and UniProt[79] (Table 2). However, valuable data are still stored in various 'in-house' or decentralized databases. Servers such as Bactibase (bacteriocins)[261], ConoServer (conotoxins)[80], CyBase (circular peptides)[262] or the PeptideAtlas (peptide spectra)[263] use their own accession codes and aim to solve pre-existing problems. The generation and deposition of data should always be a mutual goal of researchers to make sure that the information generated does not fade away as, eventually, websites will be archived, and information lost. As was the case for ArachnoServer[264], a database for spider venoms, which no longer responds to connection requests, whereas copies of its content remain on UniProt cross-referenced databases. Data deposition does not overwrite the utility of these databases, as many of them also come equipped with various tools or query-specific types of data. The best workarounds currently available are those of MiBIG[81] and the VEuPathDB project[77]. They connect data from the NCBI, the EMBL or other sources to their specific applications. Researchers are encouraged to make this the standard practice, to deposit biological sequence data with the major repositories (NCBI, EMBL, DDBJ), fetch data through their services and focus on adding layers relevant to their field onto that information. The VeuPathDB project has several resources for researchers, and MiBIG is a centralized resource used by most tools in natural product discovery. Even if the tools may be lost to the aeons, researchers are encouraged to deposit their codes on publicly available servers if tools are to be discontinued (for example, Github, Bitbucket, SourceForge). For future prospects, the Protein Data Bank (PDB)[265] has potentially the most important feature of any database, the option to deposit unpublished protein structures. This will hopefully become more standard practice where researchers can deposit unpublished experimental data under certain guidelines, as this would deal with the most significant loss of scientific data, and bring them out of the cupboards and into the Big Data landscape.

Bioinformatics, systems biology and artificial intelligence

Current bioinformatics approaches have been used successfully to identify new genes with machine learning. The machine learning methods primarily relied on hidden Markov models, support vector machines or random forest algorithms that laid the foundation for most bioinformatic approaches today[266]. Currently, deep learning algorithms[267] are becoming more frequent, as they have the advantage of being able to learn more complex features than their predecessor. Deep learning has been successfully implemented in the discovery of new genes and peptides[89], but possibly its most impressive feat is the accurate prediction of protein structures[74,268]. As the algorithms continue to improve along with access to graphics processing units to train neural networks, researchers in all fields, even with relatively little experience in programming, will be able to make use of the power of deep learning for their research. With the coming improvements from bioinformatics, bioanalytics and the omics fields, the discipline of systems biology aims to harness all levels of data it can, to understand further how each of these fields may work together[269]. Systems biology approaches have been applied in the field of metabolomics by generating genome-scale metabolic models[270]. These approaches are commonly used for production optimization, and industry relies on them to improve yields from fermentation[271,272]. The field shows promise in combining biological data for clinical applications and could facilitate the transition into personalized medicine[94].

Biological sciences are in a major transition into the big data landscape, where a lot of the focus has been on genomics, transcriptomics, proteomics and metabolomics. Peptidomics emerges as a bridge connecting proteomics and metabolomics, bridging the functions between proteins and small molecules. With the advances in deep learning and artificial intelligence, the biochemical space made available by peptides can be better exploited, and novel peptidomimetics can be developed for medicine or industry. Spanning from novel therapeutics to peptide-assisted catalysts[10], the field of peptidomics has just started to show a tiny portion of its tremendous potential.

Published online: 30 March 2023

References

1. Gruber, C. W., Muttenthaler, M. & Freissmuth, M. Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors. *Curr. Pharm. Des.* **16**, 3071–3088 (2010).
2. Dang, T. & Süssmuth, R. D. Bioactive peptide natural products as lead structures for medicinal use. *Acc. Chem. Res.* **50**, 1566–1576 (2017).
3. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discov. Today* **20**, 122–128 (2015).
4. Muttenthaler, M., King, G. F., Adams, D. J. & Alewood, P. F. Trends in peptide drug discovery. *Nat. Rev. Drug Discov.* **20**, 309–325 (2021).
This comprehensive review discusses the importance of peptides as drug leads and innovative therapeutics.
5. Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* **81**, 136–147 (2013).
6. Munch, J., Standker, L., Forssmann, W. G. & Kirchhoff, F. Discovery of modulators of HIV-1 infection from the human peptidome. *Nat. Rev. Microbiol.* **12**, 715–722 (2014).
7. Baggerman, G. et al. Peptidomics. *J. Chromatogr. B* **803**, 3–16 (2004).
8. Schrader, M., Schulz-Knappe, P. & Fricker, L. D. Historical perspective of peptidomics. *EuPA Open. Proteom.* **3**, 171–182 (2014).
9. Schulz-Knappe, P. et al. Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb. Chem. High. T Scr.* **4**, 207–217 (2001).
10. Metrano, A. J. et al. Asymmetric catalysis mediated by synthetic peptides, version 2.0: expansion of scope and mechanisms. *Chem. Rev.* **120**, 11479–11615 (2020).
This review article discusses peptide-assisted asymmetric synthesis reactions and recent advances in the field.
11. Collier, J. H. & Segura, T. Evolving the use of peptides as components of biomaterials. *Biomaterials* **32**, 4198–4204 (2011).
12. Agnieray, H., Glasson, J. L., Chen, Q., Kaur, M. & Domigan, L. J. Recent developments in sustainably sourced protein-based biomaterials. *Biochem. Soc. Trans.* **49**, 953–964 (2021).
13. Malandrino, N. & Smith, R. J. in *Principles of Endocrinology and Hormone Action* (eds Belfiore, A. & LeRoith, D.) 29–42 (Springer International, 2018).
14. Yi, J., Warunek, D. & Craft, D. Degradation and stabilization of peptide hormones in human blood specimens. *PLoS ONE* **10**, e0134427 (2015).
15. Svensson, M. et al. Heat stabilization of the tissue proteome: a new technology for improved proteomics. *J. Proteome Res.* **8**, 974–981 (2009).
16. Yang, N., Anapindi, K. D. B., Romanova, E. V., Rubakhin, S. S. & Sweedler, J. V. Improved identification and quantitation of mature endogenous peptides in the rodent hypothalamus using a rapid conductive sample heating system. *Analyst* **142**, 4476–4485 (2017).
17. Feist, P. & Hummon, A. B. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.* **16**, 3537–3563 (2015).
18. Harrison, S. T. Bacterial cell disruption: a key unit operation in the recovery of intracellular products. *Biotechnol. Adv.* **9**, 217–240 (1991).
19. Koehbach, J. et al. Cyclotide discovery in Gentianales revisited — identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. *Biopolymers* **100**, 438–452 (2013).
20. Chen, E. I., Cociorva, D., Norris, J. L. & Yates, J. R. 3rd Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **6**, 2529–2538 (2007).
21. Panuwet, P. et al. Biological matrix effects in quantitative tandem mass spectrometry-based analytical methods: advancing biomonitoring. *Crit. Rev. Anal. Chem.* **46**, 93–105 (2016).

22. Finoulst, I., Pinkse, M., Van Dongen, W. & Verhaert, P. Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. *J. Biomed. Biotechnol.* **2011**, 245291 (2011).
23. Tubaon, R. M., Haddad, P. R. & Quirino, J. P. Sample clean-up strategies for ESI mass spectrometry applications in bottom-up proteomics: trends from 2012 to 2016. *Proteomics* **17**, 1700011 (2017).
24. Sosalagere, C., Adesegun Kehinde, B. & Sharma, P. Isolation and functionalities of bioactive peptides from fruits and vegetables: a reviews. *Food Chem.* **366**, 130494 (2022).
25. Mthembu, S. N., Sharma, A., Albericio, F. & de la Torre, B. G. Breaking a couple: disulfide reducing agents. *Chembiochem* **21**, 1947–1954 (2020).
26. Hellinger, R. et al. Importance of the cyclic cystine knot structural motif for immunosuppressive effects of cyclotides. *ACS Chem. Biol.* **16**, 2373–2386 (2021).
27. Tsai, P. L., Chen, S. F. & Huang, S. Y. Mass spectrometry-based strategies for protein disulfide bond identification. *Rev. Anal. Chem.* **32**, 257–268 (2013).
28. Han, D. K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951 (2001).
29. Yao, X., Freas, A., Ramirez, J., Demirev, P. A. & Fenselau, C. Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842 (2001).
30. Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843–6852 (2003).
31. Greer, T., Lietz, C. B., Xiang, F. & Li, L. Novel isotopic *N*,*N*-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. *J. Am. Soc. Mass Spectrom.* **26**, 107–119 (2014).
32. DeSouza, L. V. et al. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J. Proteome Res.* **7**, 3525–3534 (2008).
33. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
34. Wiese, S., Reidegeld, K. A., Meyer, H. E. & Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**, 340–350 (2007).
35. Zhang, J., Wang, Y. & Li, S. Deuterium isobaric amine-reactive tags for quantitative proteomics. *Anal. Chem.* **82**, 7588–7595 (2010).
36. Atkins, N. Jr. et al. Functional peptidomics: stimulus- and time-of-day-specific peptide release in the mammalian circadian clock. *ACS Chem. Neurosci.* **9**, 2001–2008 (2018).
37. Gedela, S. & Medicherla, N. R. Chromatographic techniques for the separation of peptides: application to proteomics. *Chromatographia* **65**, 511–518 (2007).
38. Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat. Protoc.* **3**, 1709–1717 (2008).
39. Mahoney, W. C. & Hermodson, M. A. Separation of large denatured peptides by reverse phase high performance liquid chromatography. Trifluoroacetic acid as a peptide solvent. *J. Biol. Chem.* **255**, 11199–11203 (1980).
40. Yoshida, T. Peptide separation by hydrophilic-interaction chromatography: a review. *J. Biochem. Biophys. Meth.* **60**, 265–280 (2004).
41. Hillenkamp, F. & Karas, M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Meth. Enzymol.* **193**, 280–295 (1990).
42. Dreisewerd, K. The desorption process in MALDI. *Chem. Rev.* **103**, 395–426 (2003).
43. Dong, X. et al. A LC-MS/MS method to monitor the concentration of HYD-PEP06, a RGD-modified Endostar mimetic peptide in rat blood. *J. Chromatogr. B* **1092**, 296–305 (2018).
44. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
45. Ludwig, C. et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
46. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).
47. Follmann, R., Goldsmith, C. J. & Stein, W. Spatial distribution of intermingling pools of projection neurons with distinct targets: a 3D analysis of the commissural ganglia in *Cancer borealis*. *J. Comp. Neurol.* **525**, 1827–1843 (2017).
48. Mechref, Y. Use of CID/ETD mass spectrometry to analyze glycopeptides. *Curr. Protoc. Protein Sci.* **68**, 12.11.1–12.11.11 (2012).
49. Riley, N. M., Malaker, S. A., Driessen, M. D. & Bertozzi, C. R. Optimal dissociation methods differ for N-and O-glycopeptides. *J. Proteome Res.* **19**, 3286–3301 (2020).
50. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
51. Tsou, C.-C. et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–264 (2015).
52. Sinitcyn, P. et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00968-7 (2021).
53. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
54. Tyler, B. J., Rayal, G. & Castner, D. G. Multivariate analysis strategies for processing ToF-SIMS images of biomaterials. *Biomaterials* **28**, 2412–2423 (2007).
55. Eberlin, L. S. et al. Desorption electrospray ionization then MALDI mass spectrometry imaging of lipid and protein distributions in single tissue sections. *Anal. Chem.* **83**, 8366–8371 (2011).
56. Bouschen, W. & Spengler, B. Artifacts of MALDI sample preparation investigated by high-resolution scanning microprobe matrix-assisted laser desorption/ionization (SMALDI) imaging mass spectrometry. *Int. J. Mass. Spectrom.* **266**, 129–137 (2007).
57. Iakab, S.-A. et al. SALDI-MS and SERS multimodal imaging: one nanostructured substrate to rule them both. *Anal. Chem.* **94**, 2785–2793 (2022).
58. Ali, A., Baby, B., Soman, S. S. & Vijayan, R. Molecular insights into the interaction of hemorphin and its targets. *Sci. Rep.* **9**, 14747 (2019).
59. Rocha, B., Ruiz-Romero, C. & Blanco, F. J. Mass spectrometry imaging: a novel technology in rheumatology. *Nat. Rev. Rheumatol.* **13**, 52–63 (2017).
60. Ramos-Vara, J. Technical aspects of immunohistochemistry. *Vet. Pathol.* **42**, 405–426 (2005).
61. Skelley, D., Brown, L. & Besch, P. Radioimmunoassay. *Clin. Chem.* **19**, 146–186 (1973).
62. Lichtman, J. W. & Conchello, J.-A. Fluorescence microscopy. *Nat. Methods* **2**, 910–919 (2005).
63. Buchberger, A. R., DeLaney, K., Johnson, J. & Li, L. Mass spectrometry imaging: a review of emerging advancements and future insights. *Anal. Chem.* **90**, 240 (2018).
This comprehensive review discusses various aspects of MSI, spanning from sample preparation and mass spectrometry instrumentation to data analysis and diverse applications.
64. Lemaire, R. et al. Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. *J. Proteome Res.* **6**, 1295–1305 (2007).
65. Kokkat, T. J. et al. Archived formalin-fixed paraffin-embedded (FFPE) blocks: a valuable underexploited resource for extraction of DNA, RNA, and protein. *Biopreserv. Biobank* **11**, 101–106 (2013).
66. Ren, Y. et al. Reagents for isobaric labeling peptides in quantitative proteomics. *Anal. Chem.* **90**, 12366–12371 (2018).
67. Truong, J. X. et al. Removal of optimal cutting temperature (OCT) compound from embedded tissue for MALDI imaging of lipids. *Anal. Bioanal. Chem.* **413**, 2695–2708 (2021).
68. Tian, Y., Bova, G. S. & Zhang, H. Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. *Anal. Chem.* **83**, 7013–7019 (2011).
69. Bogdanow, B., Zauber, H. & Selbach, M. Systematic errors in peptide and protein identification and quantification by modified peptides. *Mol. Cell Proteom.* **15**, 2791–2801 (2016).
70. Schwartz, S. A., Reyzer, M. L. & Caprioli, R. M. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J. Mass. Spectrom.* **38**, 699–708 (2003).
71. Lemaire, R. et al. MALDI-MS direct tissue analysis of proteins: improving signal sensitivity using organic treatments. *Anal. Chem.* **78**, 7145–7153 (2006).
72. Buchberger, A. R., Sauer, C. S., Vu, N. Q., DeLaney, K. & Li, L. Temporal study of the perturbation of crustacean neuropeptides due to severe hypoxia using 4-plex reductive dimethylation. *J. Proteome Res.* **19**, 1548–1555 (2020).
73. Kaletaş, B. K. et al. Sample preparation issues for tissue imaging by imaging MS. *Proteomics* **9**, 2622–2633 (2009).
74. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
75. Leopold, J., Popkova, Y., Engel, K. M. & Schiller, J. Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids. *Biomolecules* **8**, 173 (2018).
76. DeLaney, K. et al. Mass spectrometry quantification, localization, and discovery of feeding-related neuropeptides in cancer borealis. *ACS Chem. Neurosci.* **12**, 782–798 (2021).
77. Amos, B. et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.* **50**, D898–D911 (2022).
78. Sayers, E. W. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
79. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
80. Kaas, Q., Yu, R., Jin, A. H., Dutertre, S. & Craik, D. J. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **40**, D325–D330 (2012).
81. Kautsar, S. A. et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **48**, D454–D458 (2020).
82. Leinonen, R. et al. The European Nucleotide Archive. *Nucleic Acids Res.* **39**, D28–D31 (2011).
83. Besemer, J. & Borodovsky, M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* **33**, W451–W454 (2005).
84. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
85. Hazarika, R. R. et al. ARA-PEPs: a repository of putative sORF-encoded peptides in *Arabidopsis thaliana*. *BMC Bioinformatics* **18**, 37 (2017).
86. Mooney, C., Haslam, N. J., Holton, T. A., Pollastri, G. & Shields, D. C. PeptideLocator: prediction of bioactive peptides in protein sequences. *Bioinformatics* **29**, 1120–1126 (2013).

87. Zhou, P. et al. Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics* **14**, 335 (2013).
88. Zhu, M. & Gribskov, M. MiPepid: microPeptide identification tool using machine learning. *BMC Bioinformatics* **20**, 559 (2019).
89. Zhang, Y., Jia, C., Fullwood, M. J. & Kwoh, C. K. DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. *Brief. Bioinform.* **22**, 2073–2084 (2021).
90. Lin, D. et al. Mining amphibian and insect transcriptomes for antimicrobial peptide sequences with rAMPage. *Antibiotics* **11**, 952 (2022).
91. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**, 661–673 (2008).
92. Dieckmann, M. A. et al. EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Res.* **49**, W185–W192 (2021).
93. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11**, 639–648 (2015).
94. Weber, T. & Kim, H. U. The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.* **1**, 69–79 (2016).
95. Blin, K. et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
This work presents the most significant genome mining platform for natural products, covering a wide range of compounds, and is a recommended read for anyone interested in natural product research.
96. Hannigan, G. D. et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **47**, e110 (2019).
97. Sélem-Mojica, N., Aguilar, C., Gutiérrez-García, K., Martínez-Guerrero, C. E. & Barona-Gómez, F. EvoMining reveals the origin and fate of natural product biosynthetic enzymes. *Microb. Genom.* https://doi.org/10.1099/mgen.0.000260 (2019).
98. Chevrette, M. G., Aicheler, F., Kohlbacher, O., Currie, C. R. & Medema, M. H. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across *Actinobacteria*. *Bioinformatics* **33**, 3202–3210 (2017).
This work discusses how SANDPUMA has aided NRP discovery and continues to provide valuable predictions for researchers involved in NRP research.
99. van Heel, A. J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281 (2018).
100. Ramesh, S. et al. Bioinformatics-guided expansion and discovery of graspetides. *ACS Chem. Biol.* **16**, 2787–2797 (2021).
101. Merwin, N. J. et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl Acad. Sci. USA* **117**, 371–380 (2020).
102. Schlaffner, C. N., Pirklbauer, G. J., Bender, A. & Choudhary, J. S. Fast, quantitative and variant enabled mapping of peptides to genomes. *Cell Syst.* **5**, 152–156.e4 (2017).
103. Ricart, E. et al. rBAN: retro-biosynthetic analysis of nonribosomal peptides. *J. Cheminform.* **11**, 13 (2019).
104. Kunyavskaya, O. et al. Nerpa: a tool for discovering biosynthetic gene clusters of bacterial nonribosomal peptides. *Metabolites* https://doi.org/10.3390/metabo11100693 (2021).
105. Konanov, D. N., Krivonos, D. V., Ilina, E. N. & Babenko, V. V. BioCAT: search for biosynthetic gene clusters producing nonribosomal peptides with known structure. *Comput. Struct. Biotechnol. J.* **20**, 1218–1226 (2022).
106. Grundemann, C., Koehbach, J., Huber, R. & Gruber, C. W. Do plant cyclotides have potential as immunosuppressant peptides? *J. Nat. Prod.* **75**, 167–174 (2012).
107. van Santen, J. A. et al. The Natural Products Atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.* **5**, 1824–1833 (2019).
This work discusses how the Natural Products Atlas provides valuable information, visualization and validation of discovered compounds.
108. Mohimani, H. et al. Dereplication of microbial metabolites through database search of mass spectra. *Nat. Commun.* **9**, 4035 (2018).
109. Diament, B. J. & Noble, W. S. Faster SEQUEST searching for peptide identification from tandem mass spectra. *J. Proteome Res.* **10**, 3871–3879 (2011).
110. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
111. Claesen, J., Valkenborg, D. & Burzykowski, T. De novo prediction of the elemental composition of peptides and proteins based on a single mass. *J. Mass. Spectrom.* **55**, e4367 (2020).
112. Lai, Z. et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **15**, 53–56 (2018).
113. Palmer, A. et al. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **14**, 57–60 (2017).
114. Novak, J., Skriba, A. & Havlicek, V. CycloBranch 2: molecular formula annotations applied to imzML data sets in bimodal fusion and LC-MS data files. *Anal. Chem.* **92**, 6844–6849 (2020).
115. Ricart, E., Pupin, M., Muller, M. & Lisacek, F. Automatic annotation and dereplication of tandem mass spectra of peptidic natural products. *Anal. Chem.* **92**, 15862–15871 (2020).
116. Gurevich, A. et al. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat. Microbiol.* **3**, 319–327 (2018).
117. Seidler, J., Zinn, N., Boehm, M. E. & Lehmann, W. D. De novo sequencing of peptides by MS/MS. *Proteomics* **10**, 634–649 (2010).
118. Yang, H., Chi, H., Zeng, W.-F., Zhou, W.-J. & He, S.-M. pNovo 3: precise de novo peptide sequencing using a learning-to-rank framework. *Bioinformatics* **35**, i183–i190 (2019).
119. Tran, N. H. et al. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **16**, 63–66 (2019).
120. Tran, N. H., Zhang, X., Xin, L., Shan, B. & Li, M. De novo peptide sequencing by deep learning. *Proc. Natl Acad. Sci. USA* **114**, 8247–8252 (2017).
121. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Meth. Mol. Biol.* **604**, 55–71 (2010).
122. Reiter, L. et al. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell Proteom.* **8**, 2405–2417 (2009).
123. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
124. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
125. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
126. Röst, H. L. et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **32**, 219–223 (2014).
127. Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteom.* **14**, 1400–1410 (2015).
128. Neilson, K. A. et al. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535–553 (2011).
129. Chang, C. et al. LFAQ: toward unbiased label-free absolute protein quantification by predicting peptide quantitative factors. *Anal. Chem.* **91**, 1335–1343 (2018).
130. Ishihama, Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteom.* **4**, 1265–1272 (2005).
131. Sivanich, M. K., Gu, T. J., Tabang, D. N. & Li, L. Recent advances in isobaric labeling and applications in quantitative proteomics. *Proteomics* **22**, e2100256 (2022).
This critical review article discusses isobaric labelling strategies for quantitative proteomics and peptidomics applications as well as current limitations and future outlooks.
132. Fonville, J. M. et al. Robust data processing and normalization strategy for MALDI mass spectrometric imaging. *Anal. Chem.* **84**, 1310–1319 (2012).
133. Deininger, S.-O. et al. Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. *Anal. Bioanal. Chem.* **401**, 167–181 (2011).
134. Källback, P., Shariatgorji, M., Nilsson, A. & Andrén, P. E. Novel mass spectrometry imaging software assisting labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections. *J. Proteom.* **75**, 4941–4951 (2012).
135. Shariatgorji, M. et al. Direct targeted quantitative molecular imaging of neurotransmitters in brain tissue sections. *Neuron* **84**, 697–707 (2014).
136. Lanekoff, I., Thomas, M. & Laskin, J. Shotgun approach for quantitative imaging of phospholipids using nanospray desorption electrospray ionization mass spectrometry. *Anal. Chem.* **86**, 1872–1880 (2014).
137. Hansen, H. T. & Janfelt, C. Aspects of quantitation in mass spectrometry imaging investigated on cryo-sections of spiked tissue homogenates. *Anal. Chem.* **88**, 11513–11520 (2016).
138. Robichaud, G., Garrard, K. P., Barry, J. A. & Muddiman, D. C. MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. *J. Am. Soc. Mass. Spectrom.* **24**, 718–721 (2013).
139. Alexander, J., Oliphant, A., Wilcockson, D. C. & Webster, S. G. Functional identification and characterization of the diuretic hormone 31 (DH31) signaling system in the green shore crab, *Carcinus maenas*. *Front. Neurosci.* **12**, 454 (2018).
140. Källback, P., Nilsson, A., Shariatgorji, M. & Andrén, P. E. msiQuant — quantitation software for mass spectrometry imaging enabling fast access, visualization, and analysis of large data sets. *Anal. Chem.* **88**, 4346–4353 (2016).
141. Arnison, P. G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Product. Rep.* **30**, 108–160 (2013).
This comprehensive review introduces the reader to RiPPs, from classification to biosynthesis and bioactivity.
142. Wiebach, V. et al. The anti-staphylococcal lipolanthines are ribosomally synthesized lipopeptides. *Nat. Chem. Biol.* **14**, 652–654 (2018).
This research article discusses a novel type of anti-staphylococcal RiPP, utilizing a short peptide conjugated with a lipid moiety.
143. Sussmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821 (2017).
This comprehensive review about NRPs explains biosynthesis, structures and bioactivity or NRPs.
144. Tang, S. et al. Discovery and characterization of a PKS-NRPS hybrid in *Aspergillus terreus* by genome mining. *J. Nat. Prod.* **83**, 473–480 (2020).
145. Zhang, Z., Wang, J., Wang, J., Wang, J. & Li, Y. Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome* https://doi.org/10.1186/s40168-020-00903-z (2020).

146. V, V. et al. Venom peptides — a comprehensive translational perspective in pain management. *Curr. Res. Toxicol.* **2**, 329–340 (2021).
147. King, G. F. & Hardy, M. C. Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. *Annu. Rev. Entomol.* **58**, 475–496 (2013).
148. Munawar, A., Ali, S. A., Akrem, A. & Betzel, C. Snake venom peptides: tools of biodiscovery. *Toxins* https://doi.org/10.3390/toxins10110474 (2018).
149. King, G. F. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert. Opin. Biol. Ther.* **11**, 1469–1484 (2011).
150. Dutertre, S. et al. Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat. Commun.* **5**, 3521 (2014).
This research article investigates the differences between the defensive and predatory venoms of cone snails.
151. Prashanth, J. R., Dutertre, S. & Lewis, R. J. in *Evolution of Venomous Animals and Their Toxins* Ch. 18 (ed. Malhotra, A.) 105–123 (Springer, 2017).
152. Coelho, P., Kaliontzopoulou, A., Rasko, M., Meijden, A. & Portugal, S. A 'striking' relationship: scorpion defensive behaviour and its relation to morphology and performance. *Funct. Ecol.* **31**, 1390–1404 (2017).
This work presents a fascinating investigation into the different methods of the defensive behaviours of scorpions, measuring both the speed and frequency of stings in response to stimuli.
153. Nisani, Z. & Hayes, W. K. Defensive stinging by *Parabuthus transvaalicus* scorpions: risk assessment and venom metering. *Anim. Behav.* **81**, 627–633 (2011).
154. Diesner, M., Predel, R. & Neupert, S. Neuropeptide mapping of dimmed cells of adult *Drosophila* brain. *J. Am. Soc. Mass. Spectrom.* **29**, 890–902 (2018).
155. Habenstein, J. et al. Transcriptomic, peptidomic, and mass spectrometry imaging analysis of the brain in the ant *Cataglyphis nodus*. *J. Neurochem.* **158**, 391–412 (2021).
156. Zeng, H. et al. Genomics- and peptidomics-based discovery of conserved and novel neuropeptides in the American cockroach. *J. Proteome Res.* **20**, 1217–1228 (2021).
157. El Filali, Z., Van Minnen, J., Liu, W. K., Smit, A. B. & Li, K. W. Peptidomics analysis of neuropeptides involved in copulatory behavior of the mollusk *Lymnaea stagnalis*. *J. Proteome Res.* **5**, 1611–1617 (2006).
158. Parmar, B. S. et al. Identification of non-canonical translation products in *C. elegans* using tandem mass spectrometry. *Front. Genet.* **12**, 728900 (2021).
159. Van Bael, S. et al. A *Caenorhabditis elegans* mass spectrometric resource for neuropeptidomics. *J. Am. Soc. Mass. Spectrom.* **29**, 879–889 (2018).
160. Wood, E. A. et al. Neuropeptide localization in *Lymnaea stagnalis*: from the central nervous system to subcellular compartments. *Front. Mol. Neurosci.* **14**, 670303 (2021).
161. DeLaney, K., Buchberger, A. & Li, L. Identification, quantitation, and imaging of the crustacean peptidome. *Methods Mol. Biol.* **1719**, 247–269 (2018).
162. DeLaney, K. & Li, L. Capillary electrophoresis coupled to MALDI mass spectrometry imaging with large volume sample stacking injection for improved coverage of *C. borealis* neuropeptidome. *Analyst* **145**, 61–69 (2019).
163. Liu, Y., Li, G. & Li, L. Targeted top-down mass spectrometry for the characterization and tissue-specific functional discovery of crustacean hyperglycemic hormones (CHH) and CHH precursor-related peptides in response to low pH stress. *J. Am. Soc. Mass. Spectrom.* **32**, 1352–1360 (2021).
164. Xu, L. L. et al. Major shrimp allergen peptidomics signatures and potential biomarkers of heat processing. *Food Chem.* **382**, 132567 (2022).
165. Phetsanthad, A. et al. Recent advances in mass spectrometry analysis of neuropeptides. *Mass. Spectrom. Rev.* **42**, 706–750 (2021).
166. Fujisawa, T. & Hayakawa, E. Peptide signaling in Hydra. *Int. J. Dev. Biol.* **56**, 543–550 (2012).
167. Monroe, E. B. et al. Exploring the sea urchin neuropeptide landscape by mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **29**, 923–934 (2018).
168. Takahashi, T. Neuropeptides and epitheliopeptides: structural and functional diversity in an ancestral metazoan Hydra. *Protein Pept. Lett.* **20**, 671–680 (2013).
169. Southey, B. R., Romanova, E. V., Rodriguez-Zas, S. L. & Sweedler, J. V. Bioinformatics for prohormone and neuropeptide discovery. *Methods Mol. Biol.* **1719**, 71–96 (2018).
This methodological article describes a pipeline for annotation of neuropeptide prohormones from genomic assemblies using freely available public toolsets and databases.
170. Hu, C. K. et al. Identification of prohormones and pituitary neuropeptides in the African cichlid, *Astatotilapia burtoni*. *BMC Genomics* **17**, 660 (2016).
171. Chan-Andersen, P. C., Romanova, E. V., Rubakhin, S. S. & Sweedler, J. V. Profiling 26,000 *Aplysia californica* neurons by single cell mass spectrometry reveals neuronal populations with distinct neuropeptide profiles. *J. Biol. Chem.* **298**, 102254 (2022).
This work presents an elegant mass spectrometry-based approach for robust categorization of large cell populations based on a single-cell neuropeptide profile.
172. Jiménez, C. R. et al. Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. *J. Neurosci.* **26**, 518–529 (2006).
173. Green, D. J. et al. cAMP, Ca^{2+}, pH_i, and NO regulate H-like cation channels that underlie feeding and locomotion in the predatory sea slug *Pleurobranchaea californica*. *ACS Chem. Neurosci.* **9**, 1986–1993 (2018).
174. Han, Y., Ma, B. & Zhang, K. SPIDER: software for protein identification from sequence tags with de novo sequencing error. *J. Bioinform. Comput. Biol.* **3**, 697–716 (2005).
175. Romanova, E. V., Aerts, J. T., Croushore, C. A. & Sweedler, J. V. Small-volume analysis of cell-cell signaling molecules in the brain. *Neuropsychopharmacology* **39**, 50–64 (2014).
176. Bai, L. et al. Characterization of GdFFD, a D-amino acid-containing neuropeptide that functions as an extrinsic modulator of the *Aplysia* feeding circuit. *J. Biol. Chem.* **288**, 32837–32851 (2013).
177. Checco, J. W. et al. *Aplysia* allatotropin-related peptide and its newly identified D-amino acid-containing epimer both activate a receptor and a neuronal target. *J. Biol. Chem.* **293**, 16862–16873 (2018).
178. Romanova, E. V. et al. Urotensin II in invertebrates: from structure to function in *Aplysia californica*. *PLoS ONE* **7**, e48764 (2012).
179. Zhang, G. et al. Newly identified *Aplysia* SPTR-gene family-derived peptides: localization and function. *ACS Chem. Neurosci.* **9**, 2041–2053 (2018).
180. Mast, D. H., Checco, J. W. & Sweedler, J. V. Differential post-translational amino acid isomerization found among neuropeptides in *Aplysia californica*. *ACS Chem. Biol.* **15**, 272–281 (2020).
181. Mast, D. H., Checco, J. W. & Sweedler, J. V. Advancing D-amino acid-containing peptide discovery in the metazoan. *Biochim. Biophys. Acta Proteins Proteom.* **1869**, 140553 (2021).
This review discusses the prevalence of enzyme-derived DAACPs among animals, physiological consequences of peptide isomerization and analytical methods for structural characterization/discovery of DAACPs.
182. Lambeth, T. R. & Julian, R. R. Differentiation of peptide isomers and epimers by radical-directed dissociation. *Methods Enzymol.* **626**, 67–87 (2019).
183. Mast, D. H., Liao, H. W., Romanova, E. V. & Sweedler, J. V. Analysis of peptide stereochemistry in single cells by capillary electrophoresis-trapped ion mobility spectrometry mass spectrometry. *Anal. Chem.* **93**, 6205–6213 (2021).
184. Checco, J. W. et al. Molecular and physiological characterization of a receptor for D-amino acid-containing neuropeptides. *ACS Chem. Biol.* **13**, 1343–1352 (2018).
185. Livnat, I. et al. A D-amino acid-containing neuropeptide discovery funnel. *Anal. Chem.* **88**, 11868–11876 (2016).
186. Yussif, B. M. & Checco, J. W. Evaluation of endogenous peptide stereochemistry using liquid chromatography-mass spectrometry-based spiking experiments. *Methods Enzymol.* **663**, 205–234 (2022).
187. Jiang, L. et al. A quantitative proteome map of the human body. *Cell* **183**, 269–283.e19 (2020).
188. Secher, A. et al. Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nat. Commun.* **7**, 11436 (2016).
This article introduces a comprehensive analytical workflow for large-scale mammalian peptidomics studies, detailing procedures ranging from sample preparation to data analysis.
189. Foster, S. R. et al. Discovery of human signaling systems: pairing peptides to G protein-coupled receptors. *Cell* **179**, 895–908.e21 (2019).
190. Hauser, A. S., Gloriam, D. E., Brauner-Osborne, H. & Foster, S. R. Novel approaches leading towards peptide GPCR de-orphanisation. *Br. J. Pharmacol.* **177**, 961–968 (2020).
191. Scarpa, A. Pre-scientific medicines: their extent and value. *Soc. Sci. Med. A Med. Psychol. Med. Sociol.* **15**, 317–326 (1981).
192. Pina, A. S., Hussain, A. & Roque, A. C. A. in *Ligand–Macromolecular Interactions in Drug Discovery: Methods and Protocols* (ed. Roque, A. C. A.) 3–12 (Humana, 2010).
193. Heinrich, M. Ethnobotany and its role in drug development. *Phytother. Res.* **14**, 479–488 (2000).
194. Campbell, I. B., Macdonald, S. J. F. & Procopiou, P. A. Medicinal chemistry in drug discovery in Big Pharma: past, present and future. *Drug Discov. Today* **23**, 219–234 (2018).
195. Camargo, A. C. M., Ianzer, D., Guerreiro, J. R. & Serrano, S. M. T. Bradykinin-potentiating peptides: beyond captopril. *Toxicon* **59**, 516–523 (2012).
196. Cesa-Luna, C. et al. Structural characterization of scorpion peptides and their bactericidal activity against clinical isolates of multidrug-resistant bacteria. *PLoS ONE* **14**, e0222438 (2019).
197. Jouiaei, M. et al. Ancient venom systems: a review on Cnidaria toxins. *Toxins* **7**, 2251–2271 (2015).
198. Jin, A. H. et al. Conotoxins: chemistry and biology. *Chem. Rev.* **119**, 11510–11549 (2019).
This review article on conotoxins explains the chemistry and biology behind their function by using 3D structural models, thus providing a deeper understanding of the topic.
199. McGivern, J. G. Ziconotide: a review of its pharmacology and use in the treatment of pain. *Neuropsychiatr. Dis. Treat.* **3**, 69–85 (2007).
200. Safavi-Hemami, H. et al. Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl Acad. Sci. USA* **112**, 1743–1748 (2015).
This article is interesting for researchers involved in peptide hormone research, discussing the weaponization of peptide hormones by animals.
201. Furman, B. L. The development of Byetta (exenatide) from the venom of the Gila monster as an anti-diabetic agent. *Toxicon* **59**, 464–471 (2012).
202. Muller, T. D., Bluher, M., Tschop, M. H. & DiMarchi, R. D. Anti-obesity drug discovery: advances and challenges. *Nat. Rev. Drug. Discov.* **21**, 201–223 (2022).
203. Rubinstein, E. & Keynan, Y. Vancomycin revisited — 60 years later. *Front. Public Health* https://doi.org/10.3389/fpubh.2014.00217 (2014).
204. Heidary, M. et al. Daptomycin. *J. Antimicrob. Chemother.* **73**, 1–11 (2018).
205. Felnagle, E. A. et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211 (2008).

206. Flores, C., Fouquet, G., Moura, I. C., Maciel, T. T. & Hermine, O. Lessons to learn from low-dose cyclosporin-a: a new approach for unexpected clinical applications. *Front. Immunol.* https://doi.org/10.3389/fimmu.2019.00588 (2019).
207. Additives, E. et al. Safety of nisin (E 234) as a food additive in the light of new toxicological data and the proposed extension of use. *EFSA J.* **15**, e05063 (2017).
208. Nakatsuji, T. & Gallo, R. L. Antimicrobial peptides: old molecules with new ideas. *J. Invest. Dermatol.* **132**, 887–895 (2012).
209. Lei, J. et al. The antimicrobial peptides and their potential clinical applications. *Am. J. Transl. Res.* **11**, 3919 (2012).
210. Zborovsky, L. et al. Improvement of the antimicrobial potency, pharmacokinetic and pharmacodynamic properties of albicidin by incorporation of nitrogen atoms. *Chem. Sci.* **12**, 14606–14617 (2021).
This work is an example of how medicinal chemistry can be used to improve the bioactive qualities of peptides.
211. Imai, Y. et al. A new antibiotic selectively kills Gram-negative pathogens. *Nature* **576**, 459–464 (2019).
212. Vilas Boas, L. C. P., Campos, M. L., Berlanda, R. L. A., de Carvalho Neves, N. & Franco, O. L. Antiviral peptides as promising therapeutic drugs. *Cell Mol. Life Sci.* **76**, 3525–3542 (2019).
213. Bosso, M., Ständker, L., Kirchhoff, F. & Münch, J. Exploiting the human peptidome for novel antimicrobial and anticancer agents. *Bioorg. Med. Chem.* **26**, 2719–2726 (2018).
214. Kuroki, A., Tay, J., Lee, G. H. & Yang, Y. Y. Broad-spectrum antiviral peptides and polymers. *Adv. Healthc. Mater.* **10**, e2101113 (2021).
215. Klein, J., Bascands, J.-L., Mischak, H. & Schanstra, J. P. The role of urinary peptidomics in kidney disease research. *Kidney Int.* **89**, 539–545 (2016).
216. Good, D. M. et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol. Cell. Proteom.* **9**, 2424–2437 (2010).
217. Argiles, A. et al. CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS ONE* **8**, e62837 (2013).
218. Roscioni, S. et al. A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. *Diabetologia* **56**, 259–267 (2013).
219. Nakamura, A. et al. High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* **554**, 249–254 (2018).
220. Kaya, I., Zetterberg, H., Blennow, K. & Hanrieder, J. R. Shedding light on the molecular pathology of amyloid plaques in transgenic Alzheimer's disease mice using multimodal MALDI imaging mass spectrometry. *ACS Chem. Neurosci.* **9**, 1802–1817 (2018).
221. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**, 346–366 (2019).
222. Chen, Z. et al. In-depth site-specific analysis of N-glycoproteome in human cerebrospinal fluid and glycosylation landscape changes in Alzheimer's disease. *Mol. Cell Proteom.* **20**, 100081 (2021).
223. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
224. Li, Q. et al. Site-specific glycosylation quantitation of 50 serum glycoproteins enhanced by predictive glycopeptidomics for improved disease biomarker discovery. *Anal. Chem.* **91**, 5433–5445 (2019).
225. Alim, F. Z. D. et al. Seasonal adaptations of the hypothalamo-neurohypophyseal system of the dromedary camel. *PLoS ONE* **14**, e0216679 (2019).
226. Yu, Q. et al. Targeted mass spectrometry approach enabled discovery of O-glycosylated insulin and related signaling peptides in mouse and human pancreatic islets. *Anal. Chem.* **89**, 9184–9191 (2017).
227. Anapindi, K. D. B., Romanova, E. V., Checco, J. W. & Sweedler, J. V. Mass spectrometry approaches empowering neuropeptide discovery and therapeutics. *Pharmacol. Rev.* **74**, 662–679 (2022).
This review article discusses the historical, current and future states of neuropeptidomics with mass spectrometry and their implications for therapeutic strategies in neurological disorders.
228. Tillmaand, E. G. et al. Peptidomics and secretomics of the mammalian peripheral sensory-motor system. *J. Am. Soc. Mass. Spectrom.* **26**, 2051–2061 (2015).
229. Ramachandran, S. et al. A conserved neuropeptide system links head and body motor circuits to enable adaptive behavior. *eLife* https://doi.org/10.7554/eLife.71747 (2021).
230. Van Damme, S. et al. Neuromodulatory pathways in learning and memory: lessons from invertebrates. *J. Neuroendocrinol.* **33**, e12911 (2021).
231. Greenwood, M. P. et al. The effects of aging on biosynthetic processes in the rat hypothalamic osmoregulatory neuroendocrine system. *Neurobiol. Aging* **65**, 178–191 (2018).
232. Pan, F. et al. Peptidome analysis reveals the involvement of endogenous peptides in mouse pancreatic dysfunction with aging. *J. Cell Physiol.* **234**, 14090–14099 (2019).
233. Hook, V., Lietz, C. B., Podvin, S., Cajka, T. & Fiehn, O. Diversity of neuropeptide cell–cell signaling molecules generated by proteolytic processing revealed by neuropeptidomics mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **29**, 807–816 (2018).
234. Anapindi, K. D. B. et al. PACAP and other neuropeptide targets link chronic migraine and opioid-induced hyperalgesia in mouse models. *Mol. Cell Proteom.* **18**, 2447–2458 (2019).
235. Jiang, Z. et al. Differential neuropeptidomes of dense core secretory vesicles (DCSV) produced at intravesicular and extracellular pH conditions by proteolytic processing. *ACS Chem. Neurosci.* **12**, 2385–2398 (2021).
236. Podvin, S. et al. Dysregulation of neuropeptide and tau peptide signatures in human Alzheimer's disease brain. *ACS Chem. Neurosci.* **13**, 1992–2005 (2022).
237. Al-Hasani, R. et al. In vivo detection of optically-evoked opioid peptide release. *eLife* https://doi.org/10.7554/eLife.36520 (2018).
238. Vitorino, R., Guedes, S., Costa, J. P. D. & Kasicka, V. Microfluidics for peptidomics, proteomics, and cell analysis. *Nanomaterials* https://doi.org/10.3390/nano11051118 (2021).
239. Ong, T. H., Tillmaand, E. G., Makurath, M., Rubakhin, S. S. & Sweedler, J. V. Mass spectrometry-based characterization of endogenous peptides and metabolites in small volume samples. *Biochim. Biophys. Acta* **1854**, 732–740 (2015).
240. Burger, T. Gentle introduction to the statistical foundations of false discovery rate in quantitative proteomics. *J. Proteome Res.* **17**, 12–22 (2018).
This work is a worthwhile introduction to the statistics behind FDRs, highly recommended for all researchers working in proteomics or peptidomics.
241. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7**, 40–44 (2008).
242. Korthauer, K. et al. A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* **20**, 118 (2019).
243. Kanz, C. et al. The EMBL nucleotide sequence database. *Nucleic Acids Res.* **33**, D29–D33 (2005).
244. Fukuda, A., Kodama, Y., Mashima, J., Fujisawa, T. & Ogasawara, O. DDBJ update: streamlining submission and access of human data. *Nucleic Acids Res.* **49**, D71–D75 (2021).
245. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
This work on the FAIR Guiding Principles is an essential read for all researchers as data management will become more important as data continue to be generated worldwide.
246. Pichler, K., Warner, K., Magrane, M. & UniProt, C. SPIN: submitting sequences determined at protein level to UniProt. *Curr. Protoc. Bioinformatics* **62**, e52 (2018).
247. Ternent, T. et al. How to submit MS proteomics data to ProteomeXchange via the PRIDE database. *Proteomics* **14**, 2233–2241 (2014).
248. Segerstrom, L., Gustavsson, J. & Nylander, I. Minimizing postsampling degradation of peptides by a thermal benchtop tissue stabilization method. *Biopreserv. Biobank.* **14**, 172–179 (2016).
249. Fridjonsdottir, E., Nilsson, A., Wadensten, H. & Andren, P. E. Brain tissue sample stabilization and extraction strategies for neuropeptidomics. *Methods Mol. Biol.* **1719**, 41–49 (2018).
250. Stingl, C., Soderquist, M., Karlsson, O., Boren, M. & Luider, T. M. Uncovering effects of ex vivo protease activity during proteomics and peptidomics sample extraction in rat brain tissue by oxygen-18 labeling. *J. Proteome Res.* **13**, 2807–2817 (2014).
251. Katz, M., Hover, B. M. & Brady, S. F. Culture-independent discovery of natural products from soil metagenomes. *J. Ind. Microbiol. Biotechnol.* **43**, 129–141 (2016).
252. Reher, R. et al. Native metabolomics identifies the rivulariapeptolide family of protease inhibitors. *Nat. Commun.* **13**, 4619 (2022).
253. Mills, R. H. et al. Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* **7**, 262–276 (2022).
254. Hellinger, R. et al. Peptidomics of circular cysteine-rich plant peptides: analysis of the diversity of cyclotides from viola tricolor by transcriptome and proteome mining. *J. Proteome Res.* **14**, 4851–4862 (2015).
255. Haynes, W. A., Tomczak, A. & Khatri, P. Gene annotation bias impedes biomedical research. *Sci. Rep.* **8**, 1362 (2018).
256. Flissi, A. et al. Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res.* **48**, D465–D469 (2020).
257. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
258. Saldivar-Gonzalez, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L. & Plisson, F. Natural product drug discovery in the artificial intelligence era. *Chem. Sci.* **13**, 1526–1546 (2022).
259. Mohimani, H. et al. Dereplication of peptidic natural products through database search of mass spectra. *Nat. Chem. Biol.* **13**, 30–37 (2017).
260. Jeanne Dit Fouque, K. et al. Fast and effective ion mobility-mass spectrometry separation of D-amino-acid-containing peptides. *Anal. Chem.* **89**, 11787–11794 (2017).
261. Hammami, R., Zouhir, A., Le Lay, C., Ben Hamida, J. & Fliss, I. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol.* **10**, 22 (2010).
262. Wang, C. K., Kaas, Q., Chiche, L. & Craik, D. J. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* **36**, D206–D210 (2008).
263. Deutsch, E. W. The PeptideAtlas Project. *Methods Mol. Biol.* **604**, 285–296 (2010).
264. Pineda, S. S. et al. ArachnoServer 3.0: an online resource for automated discovery, analysis and annotation of spider toxins. *Bioinformatics* **34**, 1074–1076 (2018).
265. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).
266. Larranaga, P. et al. Machine learning in bioinformatics. *Brief. Bioinform.* **7**, 86–112 (2006).
This interesting review discusses the machine learning methods that got bioinformatics to where it is today.
267. Min, S., Lee, B. & Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **18**, 851–869 (2017).
This article describes the use, applications and architecture of deep learning networks, providing the readers with insight into the direction that bioinformatics is heading in the next decade.
268. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
269. Breitling, R. What is systems biology? *Front. Physiol.* **1**, 9 (2010).

270. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
271. Mitra, S., Dhar, R. & Sen, R. Designer bacterial cell factories for improved production of commercially valuable non-ribosomal peptides. *Biotechnol. Adv.* **60**, 108023 (2022).
272. Helmy, M., Smith, D. & Selvarajoo, K. Systems biology approaches integrated with artificial intelligence for optimized metabolic engineering. *Metab. Eng. Commun.* **11**, e00149 (2020).

Acknowledgements
Work in the laboratory of C.W.G. has been supported by the Austrian Science Fund (FWF) through projects P32109 and ZK 81B. The work of A.S. and R.D.S. was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy — EXC 2008-390540038 (UniSysCat) and RTG 2473 'Bioactive Peptides'. The work of L.L. is supported by the National Science Foundation (NSF) (CHE-2108223) and National Institutes of Health (NIH) through grants (R01DK071801, R01 AG078794 and RF1AG052324). The work of J.V.S. is supported by the National Institute on Drug Abuse under Award No. P30DA018310 and the National Institute of Neurological Disorders and Stroke (NINDS) through R01NS031609.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43586-023-00205-2.

Peer review information *Nature Reviews Methods Primers* thanks Vivian Hook, Dong-Woo Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Related links
antiSMASH: https://antismash.secondarymetabolites.org
DeepBGC: https://github.com/Merck/deepbgc
DeepRiPP: http://deepripp.magarveylab.ca
DEREPLICATOR+: https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp
Dictionary of Natural Products: https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml
High Definition Imaging: https://www.waters.com/waters/en_US/High-Definition-Imaging-(HDI)-Software/nav.htm?cid=134833914&locale=en_US
ImageQuest: https://www.thermofisher.com/order/catalog/product/10137985
MS-FINDER: http://prime.psc.riken.jp/compms/msfinder/main.html
MSiReader: https://msireader.com/
msiQuant: https://ms-imaging.org/paquan/
NCBI: https://www.ncbi.nlm.nih.gov
SANDPUMA: https://bitbucket.org/chevrm/sandpuma/src/master/
SCiLS Lab: https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/scils-lab.html
UniProt: https://www.uniprot.org