

Review

Not peer-reviewed version

A Comprehensive Review of Plant Peptide Databases for Peptidomic and Therapeutic Research

[Jahangir Alom](#) , [Sony Kumari](#) , Ujwal P. *

Posted Date: 24 April 2026

doi: 10.20944/preprints202604.1610.v1

Keywords: plant peptide databases; plant peptidomics; antimicrobial peptides; bioactive peptides; functional annotation; data curation; sustainable agriculture



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

A Comprehensive Review of Plant Peptide Databases for Peptidomic and Therapeutic Research

Jahangir Alom ¹, Sony Kumari ² and Ujwal P. ^{1,*}

¹ Department of Biotechnology Engineering, NMAM Institute of Technology, Nitte, NITTE (Deemed to be University, Nitte - 574110, Karnataka, India

² School of Biological Sciences, University of Science and Technology Meghalaya-793101

* Correspondence: ujwal.p@nitte.edu.in

Abstract

The bioactive plant peptides represent a significant and yet largely underexploited resource with huge potential not only for basic plant science but also for various biotechnological applications, including pharmaceutical and agrochemical development. bioactive plant peptides represent a significant and yet largely underexploited resource with huge potential not only for basic plant science but also for various biotechnological applications, including pharmaceutical and agrochemical development. This fast-expanding research area of plant peptidomics demands the creation and continuous updating of dedicated databases that facilitate data integration of heterogeneous nature and enable efficient knowledge discovery. The most representative databases, like PlantPepDB and PhytAMP, but also recent multi-purpose databases like MFPPDB, are reviewed here in terms of data sources used-literature and public repositories-manual curation extent, functional classification-such as therapeutic, defense, and inhibitory-and the availability of relevant metadata on physicochemical properties and structure. While databases focused on specific bioactivities of plant peptides offer high-quality, focused data, broader repositories are crucial for discovering multifunctional peptides and structure-activity relationships. The refinement and integration of these databases, alongside advanced bioinformatics tools, remain essential for overcoming these hurdles. These resources stand to facilitate innovation in ways that will continue to illuminate insights into the molecular function of plants and allow the successful harnessing of plant peptides toward human health improvements and sustainable agriculture. This review briefly introduces the progress of plant peptide research, presents an overview of plant peptide studies, and provides a comprehensive analysis of existing plant peptide databases, evaluating their scope, content, and utility. We anticipate that this work will bridge the gap between peptide discovery and the development of next-generation plant peptide databases.

Keywords: plant peptide databases; plant peptidomics; antimicrobial peptides; bioactive peptides; functional annotation; data curation; sustainable agriculture

1. Introduction

Plant peptides represent a promising avenue for botanical and biotechnological research aimed at addressing challenges in agriculture, environmental sustainability, human health, and industrial applications. These peptides play crucial roles in plant growth, development, and defence mechanisms, often acting as signalling molecules that regulate various physiological processes [1]. Plant peptides are involved in fundamental processes such as seed germination, root growth, flowering, and fruit development [2,3]. They regulate cell division, differentiation, and tissue patterning, ensuring proper plant structure and function [4]. Plants produce peptides in response to environmental stresses such as drought, salinity, pathogens, and temperature fluctuations. These peptides help plants adapt and survive under adverse conditions by triggering appropriate defence mechanisms [5]. Many plant peptides act as signalling molecules that mediate communication

between cells, tissues, and organs. They regulate gene expression, hormone levels, and metabolic pathways, orchestrating complex physiological responses. Some peptides facilitate symbiotic relationships with beneficial microbes, enhancing nutrient uptake, disease resistance, and overall plant health [6].

Beyond their functions in plants, research has unveiled a plethora of bioactive properties that can positively impact human health and applications in food. They have been proven to have various functional activities such as antioxidant, antihypertensive, immunomodulatory, antimicrobial, anti-inflammatory, antidiabetic, antithrombotic, and so on [7]. The diverse biological activities of plant peptides are extensively documented in the Plant Peptide Database, which serves as a vital resource for researchers exploring the diverse functions of these plant-derived peptides [8]. In this review, briefly introduces the progress of plant peptide research and provide valuable information about the plant peptide databases, computational tools as resources to fosters a deeper understanding of plant peptide biology and its applications.

2. A Categorized Overview of the Major Plant Peptide Databases

2.1. Focus on Model Plant Species

The development of plant peptide databases reflects the prioritization of certain model species due to their well-annotated genomes and extensive research. The model plant species in plant peptide research is crucial, as these species serve as foundational systems for understanding the complex signaling mechanisms and defense responses in plants [9]. Approximately 13,000 genes encoding these peptides have been identified in model species, highlighting their prevalence and potential roles in plant defense mechanisms. The model plant species have been extensively utilized due to their well-characterized genomes, rapid life cycles, and ease of genetic manipulation, facilitating the exploration of peptide functions and interactions across diverse plant species and serve as comprehensive resource for the discovery, production, and application of bioactive peptides [10].

2.1.1. Arabidopsis-Centric Databases

A substantial number of plant peptide databases focus on *Arabidopsis thaliana*, a model plant species with a highly detailed and well-annotated genome [11]. This concentration on Arabidopsis is strategically advantageous due to the ease of protein identification facilitated by the comprehensive genomic information. Key databases within this category include the Plant Protein DataBase (PPDB), the SUBcellular Arabidopsis database (SUBA), and the MASCOP Gator [12–14]. The rationale for this focus on Arabidopsis as a central model organism for plant proteomics is extensively detailed in the work of Weckwerth et al. [15], who highlight the model's significance in facilitating large-scale comparative studies and data standardization across the research community. The availability of a complete genome sequence for Arabidopsis provides an unparalleled resource for protein identification using mass spectrometry (MS) analysis of plant samples [16]. This has significantly advanced proteomic studies in plants and contributed to the development of many Arabidopsis-focused databases.

2.1.2. Databases for Other Model Species

While Arabidopsis remains central, the expanding landscape of plant peptidomics also encompasses databases dedicated to other significant plant species, reflecting the growing availability of genomic information for these organisms [17]. Notable examples include resources focused on rice (*Oryza sativa*) and soybean (*Glycine max*) [18]. Model plants like *Zea mays* (corn) and *Nicotiana benthamiana* are extensively studied for their peptide profiles. Research has identified peptides with therapeutic potential, such as those from corn that may inhibit enzymes related to chronic diseases. *Nicotiana benthamiana* has been optimized for high-yield production of AMPs, demonstrating scalability for therapeutic applications [19,20]. These species are of substantial

agricultural importance, and their inclusion in dedicated databases reflects the increasing need for detailed proteomic information relevant to crop improvement and stress response studies. The development of these databases is directly linked to the progress in genomic sequencing technologies, which have made it possible to obtain high-quality genomic data for a wider range of plant species [21]. The availability of this genomic information allows for more accurate protein identification and annotation, leading to the creation of more comprehensive and reliable databases.

2.2. Data Type and Peptide Classes

Plant peptide databases exhibit diversity in their focus, encompassing a range of data types and specific peptide classes. Plant peptide databases can be categorized based on the type of data they provide, which includes experimentally validated peptides, predicted peptides, and non-coding RNA-encoded peptides and based on the classes of peptides they contain, particularly focusing on their functional and structural characteristics [22]. These databases serve as essential resources for researchers exploring the therapeutic potential of plant-derived peptides. Each database serves distinct research needs and offers unique features.

2.2.1. Proteome Databases

Several databases prioritize comprehensive proteomic data, including protein identification, subcellular localization, and post-translational modifications (PTMs) [23]. A prime example is PPDB [12], which integrates various data types, including MS-derived information and literature-curated data, to provide a detailed overview of the Arabidopsis and maize proteomes. Sun et al. [12] extensively detail the methodologies employed in data acquisition and curation for PPDB, emphasizing the use of mass spectrometry for cell type-specific proteomes and subcellular proteomes. Similarly, Zybaïlov et al. [24] present a large-scale analysis of the Arabidopsis chloroplast proteome, highlighting the identification of over 1300 proteins and the determination of their abundance using spectral counting. The detailed methodology employed in this study provides valuable insight into the techniques used for data acquisition and analysis in proteome databases.

2.2.2. Phosphoproteome Databases

Another important category focuses on phosphoproteome data, specifically detailing phosphorylation events and their roles in signaling pathways [25]. PhosPhAt is a prominent example, providing a comprehensive resource for phosphorylation sites in Arabidopsis thaliana. This database not only compiles experimentally identified phosphorylation sites but also incorporates a plant-specific phosphorylation site predictor. The methods used for identifying and predicting phosphorylation sites are described in detail in these studies, including the use of mass spectrometry and machine learning algorithms [26]. The Plant Protein Phosphorylation Database (P³DB) provide comprehensive repository for plant protein phosphorylation data, facilitating the identification of functionally conserved phosphorylation sites across multiple plant species [27]. The importance of phosphoproteome databases lies in their ability to provide insights into the complex regulatory mechanisms governing plant cellular processes [28].

2.2.3. Specialized Peptide Databases

Beyond broad proteome and phosphoproteome data, specialized databases concentrate on particular peptide types, recognizing the diverse biological roles of plant peptides [29]. Examples include resources dedicated to antimicrobial peptides (AMPs) and bioactive peptides. The **PhytAMP & C-PAmP** dedicated to antimicrobial peptides, it categorizes peptides into families such as defensins, thionins and employs a classification algorithm to identify potential antimicrobial peptides. These databases provides taxonomic and microbiological data, facilitating the study of these peptides as alternatives to antibiotics [30,31]. **MFPPDB** integrates data from various functional peptide databases and includes over 1.4 million peptides with multiple therapeutic functions, categorized

into 41 features such as anti-cancer and anti-viral activities. These specialized databases are crucial because they focus on specific functional classes of peptides, allowing researchers to easily access information relevant to their specific research interests [32]. Aguilera-Mendoza et al. address the issue of redundancy in AMP databases, proposing a new non-redundant database to improve data quality and accessibility. The study highlights the challenges associated with managing and integrating data from multiple sources [33]. Other studies focus on the applications of AMPs in various fields, including agriculture, pharmaceuticals, and food preservation. Similarly, studies focusing on bioactive peptides highlight their potential applications in human health and disease prevention [34,35]. These resources provide valuable information on the diverse biological activities and potential applications of specific types of plant peptides.

2.3. Data Integration and Aggregation

The complexity of plant peptide systems necessitates strategies for data integration and aggregation to provide a more holistic view. The integration and aggregation of plant peptide data have led to the development of several comprehensive databases that enhance research in plant biology and therapeutic applications.

2.3.1. Integrative Databases

Some databases adopt an integrative approach, combining diverse omics data for a more comprehensive understanding of plant biology [36,37]. GabiPD exemplifies this approach, serving as a gateway for the German plant community to consolidate various research programs into a single resource [38]. The **ncPlantDB** integrates non-coding RNAs (ncRNAs) and their encoded peptides (ncPEPs) across 43 plant species, combining data from established databases and literature mining, enhancing research on plant ncRNA functions and interactions [39]. The **PGSB/MIPS PlantsDB** integrated search functionality across multiple plant genome databases for comparative genomics studies [40]. The integrative nature of such databases allows researchers to explore the interplay between different biological levels, such as genomics, transcriptomics, and proteomics [41]. Das et al. (2018) discuss the challenges and opportunities associated with integrating various omics data types, emphasizing the need for standardized data formats and robust data integration methods. The ability to access multiple data types within a single database greatly enhances the depth and scope of research [42].

2.3.2. Aggregative Portals

To improve accessibility and data visualization, aggregative portals such as MASCP Gator collect data from multiple sources [43]. These portals provide a unified interface for accessing and visualizing proteomic data from various databases, streamlining the research process and reducing the need to navigate multiple individual databases. The MASCP Gator, showcasing its value in providing a centralized resource for Arabidopsis proteomic data [44]. The PlantAFP database aggregates experimentally verified antifungal plant peptides from research articles, patents, public databases and provide different parameter of the collected peptides [45]. The development of such portals reflects the growing need for efficient data management and integration within the plant peptidomics community.

3. A Comparative Analysis of Database Features and Functionalities

The comparative analysis of plant peptide databases reveals significant differences in their features and functionalities, catering to diverse research needs in the field of plant peptidomics. Each database offers unique attributes that enhance the accessibility and utility of plant-derived peptides for research. This section compares the features and functionalities of various plant peptide databases, focusing on data search and retrieval, data annotation and curation, and data accessibility and user-friendliness [46,47].

3.1. Data Search and Retrieval

Data search and retrieval from plant peptide databases involve accessing comprehensive repositories that catalog plant-derived peptides with various functionalities. Efficient data retrieval is crucial for the usability of any database [48]. The following aspects are important for comparison:

3.1.1. Search Capabilities

Databases vary in their search capabilities. Some offer basic keyword searches, while others provide more advanced options, such as sequence homology searches, searches based on post-translational modifications (PTMs), and searches based on subcellular localization [49]. The sophistication of search capabilities significantly impacts the efficiency of data retrieval. A database with limited search options may require extensive manual filtering, which can be time-consuming and inefficient. Advanced search features allow for more precise and targeted searches, saving significant time and effort [50].

3.1.2. Data Filtering and Sorting

The ability to filter and sort search results is crucial for managing large datasets. Databases offer diverse filtering options, such as filtering by species, tissue type, PTMs, or other relevant parameters [51,53]. Sorting options allow users to order results by various criteria, such as protein abundance, molecular weight, or pI value [52]. The availability of sophisticated filtering and sorting options greatly enhances the usability of a database, particularly when dealing with large and complex datasets. It enables users to quickly narrow down their searches to the most relevant results, saving time and effort [54].

3.1.3. Data Visualization

Effective data visualization tools are essential for exploring database content. Some databases incorporate graphical representations of data, such as phylogenetic trees, protein interaction networks, or heatmaps [55]. These visualizations facilitate the identification of patterns and trends in the data, providing valuable insights that might be missed through simple textual data presentation. The quality and type of visualization tools provided can significantly impact the user experience and the ability to extract meaningful information from the database [56].

3.2. Data Annotation and Curation

The Plant peptide databases significantly enhance the accessibility and utility of plant peptide information, challenges remain in ensuring the accuracy and completeness of annotations, particularly for unannotated peptides, which may still hold undiscovered therapeutic potential and biological activities. Data annotation and curation in plant peptide databases involve utilizing resources like G-PTM-D for PTM information and databases such as Ensemble, RefSeq, and UniProt to enhance peptide identification and provide comprehensive protein information for diverse applications in proteomics. The quality of data annotation and curation is paramount for ensuring the accuracy and reliability of database information [57–59].

3.2.1. Annotation Quality

The completeness and accuracy of annotations vary significantly across databases. High-quality annotations include detailed information about protein identification, function, subcellular localization, PTMs, and other relevant properties. Inconsistent or incomplete annotations can lead to errors in data interpretation and hinder research progress. The quality of annotations reflects the effort invested in data curation and validation [60–62].

3.2.2. Data Validation

Data validation is crucial for ensuring the accuracy and reliability of database information. Databases employ various validation strategies, such as manual review of data, comparison with other databases, and the use of statistical methods to identify and remove errors. Rigorous data validation procedures are essential for maintaining the integrity and trustworthiness of the database. The level of data validation reflects the commitment of the database maintainers to data quality and accuracy [63,64].

3.2.3. Data Update Frequency

Regular updates are critical for keeping databases current with the latest research findings. Databases vary in their update frequency, with some updating regularly, while others update less frequently. The frequency of updates impacts the timeliness and relevance of the information provided. A database with infrequent updates may contain out dated or incomplete information, hindering research progress [65,67].

3.3. Data Accessibility and User Friendliness

The accessibility and user-friendliness of plant peptide databases are critical for researchers aiming to explore the various criteria of plant-derived peptides including biological activities and therapeutic potential. Accessibility and user-friendliness are key factors in determining the overall usability of a database [67].

3.3.1. Web Interface Design

A well-designed web interface is essential for easy navigation and data access. User-friendly interfaces incorporate intuitive navigation menus, clear search options, and effective data visualization tools. A poorly designed interface can be frustrating and hinder data access, leading to inefficient use of the database [68,69].

3.3.2. Data Download Options

The availability of various data download formats is crucial for enabling researchers to integrate database information into their own analyses. Databases should provide options for downloading data in common formats, such as FASTA, CSV, or XML. The flexibility of download options enhances the utility of the database by allowing integration with other bioinformatics tools and workflows [70].

3.3.3. Documentation and Support

Comprehensive documentation and technical support are crucial for helping users effectively utilize database resources. High-quality documentation provides detailed explanations of database features, search options, and data formats. Accessible technical support channels, such as email or FAQs, can assist users in troubleshooting problems and resolving queries. The availability of comprehensive documentation and responsive technical support enhances the user experience and promotes effective use of the database [71].

Table 1. Databases for Plant Peptides.

SL. No.	Peptide Database	Description	URL
1	Plant Genome Database (PGDB)	Provides genomic data on plants, including peptide sequences derived from transcriptomics and proteomics studies.	https://www.plantgenome.org/

2	Plant Peptide Database (PPD)	One of the largest plant peptide databases, containing both known and predicted peptides across various plant species.	https://plantpeptide.science.ddbj.nig.ac.jp
3	PeptideDB	A focused database on antimicrobial peptides (AMPs) in plants, crucial for plant defense mechanisms.	https://www.peptidedb.org
4	PLANTCYC	Integrates peptide data with metabolic pathways, providing insights into the biochemistry and functions of plant peptides.	https://plantcyc.org
5	AntiPlantPePdb	A database dedicated to plant peptides with antimicrobial properties, often involved in plant-pathogen interactions.	http://antiplantpepd.jnu.ac.in
6	PlantGDB	A major resource for plant gene data, including those coding for peptides. It covers a wide variety of species.	http://www.plantgdb.org
7	PlantAMP Database	A focused database on antimicrobial peptides (AMPs) in plants, crucial for plant defense mechanisms.	https://www.cbs.dtu.dk/services/PlantAMP
8	PepBind	Focuses on plant peptides that interact with binding proteins, playing roles in signaling and regulation.	http://pepbinding.org
9	Peptaibol Database	A database for peptaibols, a class of plant peptides with antimicrobial properties.	http://www.peptaibol.org
10	PhyPep	A comprehensive database cataloging peptides from various plant species with evolutionary insights and functional data.	https://www.phypep.org
11	Arabidopsis Peptide Database (AtPePDB)	A specialized database for peptides from the model plant <i>Arabidopsis thaliana</i> .	https://www.arabidopsis-peptide-db.org
12	Soybean Peptide Database (SoyPepDB)	Focuses on peptides from <i>Glycine max</i> , including those involved in plant stress response and defense.	https://www.soypepd.org
13	LegumePePdb	This database focuses on peptides from legumes, an important group of plants for agriculture.	http://legumepedb.jnu.ac.in
14	MASCP Gator	MASCP Gator offers features like peptide search, peptide hover, enhanced track navigation, and data visualization tracks for various aspects of peptide and protein information related to <i>Arabidopsis thaliana</i> .	http://gator.masc-proteomics.org/
15	PhosPhAt 4.0	Arabidopsis Protein Phosphorylation Site Database (PhosPhAt 4.0) predict phosphorylation site and insight about peptide properties.	https://phosphat.uni-hohenheim.de/

16	PhytAMP	PhytAMP database provides valuable information on antimicrobial plant peptides.	http://phytamp.pfba-lab-tun.org/main.php
17	C-PAmP	C-PAmP database of computationally predicted plant antimicrobial peptides.	http://bioserver-2.bioacademy.gr/Bioserver/C-PAmP/
18	MFPPDB	A comprehensive multi-functional plant peptide database with 1,482,409 peptides from 121 plant	http://124.223.195.214:9188/mfppdb/index
19	GabiPD	GabiPD provide Proteomics data, PTM and functional plant protein.	https://www.gabipd.org/
20	ncPlantDB	ncPlantDB is a non-coding RNA database for plants that focuses on cataloging non-coding RNAs (ncRNAs, ncRNAs are important for the regulation of peptide and protein functions in plants.	http://www.ncplantdb.com/
21	PGSB/MIPS PlantsDB	The PlantsDB at PGSB/MIPS serves as a resource for genomic, transcriptomic, proteomic data as well as includes peptide sequences and protein-coding genes. of a wide range of plant species.	https://plants.ensembl.org
22	PlantAFP	It is a specialized database dedicated to Antifungal Peptides (AFPs) in plants.	http://bioinformatics.cimap.res.in/sharma/PlantAFP/ .
23	BIOPEP-UWM	This bioinformatics resource designed to analyze and predict bioactive peptides.	https://biochemia.uwm.edu.pl/biopep/start_biopep.php
24	Phytoserf	Phytoserf is a database that integrates information on small secreted proteins (SSPs) and peptides from plants including SSP sequences, annotations, and functional classifications.	http://phytoserf.com/
25	AHPD	Arabidopsis Hormone Peptide Database (AHPD) is a specialized database that focuses on peptide hormones in Arabidopsis thaliana.	https://www.ahpd.uni-bayreuth.de

4. Methodological Considerations in Plant Peptide Database Construction

The construction and maintenance of plant peptide databases involve intricate methodological considerations related to data acquisition, curation, prediction and validation.

4.1. Data Acquisition Techniques

For identification of plant peptide sequences and generation of data for plant peptide databases relies heavily on advancements in mass spectrometry and bioinformatics. The methods can be categorized into bioinformatic identification, extraction techniques, and structural elucidation, each contributing to the comprehensive understanding of plant peptides [72].

4.1.1. Mass Spectrometry

Mass spectrometry (MS) is the cornerstone of plant peptidomics, playing a crucial role in identifying and quantifying peptides [73]. Various MS techniques are employed, each with its

strengths and limitations. For example, tandem mass spectrometry (MS/MS) is widely used for peptide sequencing, while label-free quantification methods, such as spectral counting or extracted ion chromatogram (XIC) analysis, are used for estimating protein abundance [74]. The choice of MS technique significantly impacts the quality and quantity of data obtained. Advances in MS technology, such as the development of high-resolution mass spectrometers and improved data acquisition methods, have significantly enhanced the sensitivity and accuracy of peptide identification and quantification [75].

Capillary electrophoresis (CE) coupled with mass spectrometry (MS)/ electrospray ionization mass spectrometry (ESI-MS) enables the identification and analysis of plant peptides, offering high resolution and detailed amino acid sequencing and compositional analysis of peptides, crucial for understanding peptides structure and function in biological systems [76,77].

4.1.2. High-Performance Liquid Chromatography (HPLC)

The HPLC is essential for the separation, isolation, and structural determination of peptides due to its high speed, resolution, and sensitivity. HPLC techniques, particularly reversed-phase chromatography, are widely utilized for analyzing complex peptide mixtures, enabling researchers to effectively isolate peptides from various sources and purify them for further study [78]. The following sections detail the key aspects of HPLC's role in peptide analysis. **Reversed-Phase HPLC** employs a hydrophobic stationary phase, allowing for the effective separation of peptides based on their hydrophobicity through gradient elution with organic solvents [79]. **Size-Exclusion and Ion-Exchange Chromatography** are also employed for peptide separation, providing additional methods to analyze peptide mixtures based on size and charge [80]. The core-shell particle technology have enhanced separation efficiency, leading to faster and more reliable results in complex samples analysis from biological and life sciences [81]. While HPLC is a powerful tool for peptide analysis, it is important to consider alternative methods such as capillary electrophoresis, which can also provide effective separation and characterization of peptides, particularly in specific contexts where HPLC may be less efficient [82].

4.1.3. Bioinformatics Tools

Bioinformatics tools are essential for processing and analyzing the vast amounts of data generated by MS, HPLC, NMR experiments [83,91]. These tools are used for peptide identification by searching against protein databases, for quantifying peptide and protein abundance, and for annotating identified peptides with functional information. The selection and application of bioinformatics tools significantly influence the accuracy and efficiency of data analysis [83,84]. The development of sophisticated bioinformatics algorithms and software has significantly improved the speed and accuracy of data analysis in plant peptidomics [85]. I-TASSER (Iterative Threading ASSEmblY Refinement) is a widely used tool for predicting the 3D structure of proteins and peptides. It combines threading, assembly, and refinement techniques to generate accurate structural models [86]. Similarly, AlphaFold, powered by artificial intelligence, has revolutionized protein structure prediction. It has been successfully applied to predict the structures of plant proteins, including those from crop plants like soybean, where 65% of UniProt sequences have predicted models [87]. Molecular docking software, such as AutoDock and Glide, is used to predict the binding of plant peptides to target proteins. This is particularly useful for understanding the mechanism of action of biological activities of plant peptides and their potential as therapeutics candidates and Molecular dynamics simulations are also employed to analyze the binding affinity of peptide to target biomolecules, aiding in the design of peptides with improved bioactivity [88,89]. The BIOPEP-UWM database and other specific plant peptide databases provide additional resources for studying plant peptides, include information on peptide sequences, functional motifs, and interactions with biomolecules [90]. The comprehensive overview of the software tools and databases used in plant proteomics, highlighting the importance of bioinformatics in addressing critical issues in data

analysis and visualization. The continual development of open-source and community-developed tools has greatly benefited the plant proteomics community [91].

4.2. Data Curation and Validation

Data curation and validation are crucial for ensuring the quality and reliability of plant peptide databases.

4.2.1. Data Cleaning and Filtering

Raw data from MS experiments often contains errors and inconsistencies that need to be addressed during data cleaning and filtering. This process involves removing low-quality spectra, identifying and correcting errors in peptide identification, and removing redundant or duplicate entries. The rigor of data cleaning and filtering procedures directly impacts the accuracy and reliability of the database. Careful data cleaning is crucial to ensure that the database contains only high-quality, reliable information [92,93].

4.2.2. Data Validation Strategies

Data validation involves comparing data from different sources, using statistical methods to detect errors, and manually reviewing data to ensure accuracy. Multiple validation strategies are often employed to ensure data reliability. The choice of validation strategies depends on the specific data type and the goals of the database. Databases with rigorous data validation procedures are more likely to contain accurate and reliable information [94].

4.2.3. Quality Control Metrics

Quality control metrics are used to assess the quality and reliability of database data. These metrics include false discovery rates (FDRs), the number of unique peptide identifications, and the completeness of annotations. The use of quality control metrics enables the assessment of data quality and identification of areas requiring improvement. The selection and application of quality control metrics are crucial for maintaining the overall quality and trustworthiness of the database [95,96].

5. Applications of Plant Peptide Databases in Research

Plant peptide databases serve as indispensable resources for a wide range of research applications. Hence databases compile extensive information on plant-derived peptides, facilitating the exploration of their diverse biological functions and potential applications [48]. The integration of detailed physicochemical properties, functional annotations, and interaction networks in these databases enhances their utility for researchers [48,97].

5.1. Discovery of Novel Peptides

Databases facilitate the identification of novel peptides through genome mining and omics data integration.

Genome mining: Plant peptide databases can be used to identify novel peptide sequences directly from plant genomes [98,99]. This involves using bioinformatics tools to search for open reading frames (ORFs) encoding peptides, followed by analysis of the predicted peptide sequences for features such as signal peptides, transmembrane domains, and conserved motifs. Porto et al. demonstrate the use of *in silico* methods to identify novel hevein-like peptide precursors from various plant species, highlighting the potential of genome mining for discovering novel peptides. The study uses a pattern-based approach to identify potential peptide precursors within sequence databases [100].

Transcriptome and proteome analysis: Integrating database information with transcriptomic and proteomic data allows for the identification of novel peptides expressed under specific conditions

[98]. This approach combines information on gene expression levels with the identification of peptides from protein extracts, providing a more comprehensive view of the peptide landscape. Hellinger et al. used a combination of transcriptome and proteome mining to characterize the cyclotide peptidome of *Viola tricolor*, demonstrating the power of integrating multiple omics datasets. The study highlights the use of mass spectrometry and bioinformatics tools for identifying and characterizing peptides, demonstrating the importance of integrating multiple data sources for a more comprehensive understanding of the peptide landscape [98].

5.2. Functional Annotation of Peptides

Databases are instrumental in annotating peptide function through sequence homology searches, phylogenetic analysis, and predictive modelling [101].

Sequence homology searches: Identifying homologous peptides with known functions helps annotate the function of newly discovered peptides. Sequence similarity often indicates functional similarity, providing a starting point for functional annotation. Databases facilitate these searches by providing tools to compare peptide sequences against known protein sequences [102].

Phylogenetic analysis: Phylogenetic analysis of peptide sequences can provide insights into the evolutionary relationships between peptides and their potential functions [103]. This approach involves constructing phylogenetic trees based on peptide sequences and comparing the phylogenetic relationships of peptides with their functional annotations. Strabala et al. used phylogenetic analysis to study the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and CLE-LIKE signal peptide genes in the Pinophyta, demonstrating the utility of this approach for understanding peptide evolution and function. The study highlights the importance of phylogenetic analysis for inferring the functional roles of peptides [103].

5.3. Predictive Modeling

Predictive models can be employed to infer peptide functions based on sequence features or other characteristics [104]. These models can be trained on datasets of peptides with known functions and used to predict the functions of novel peptides. Kaur et al. developed a prediction server for identifying peptide hormones using a combination of machine learning and similarity-based methods, showcasing the potential of predictive modeling for annotating peptide function. The study highlights the use of various machine learning algorithms for developing a prediction model and the integration of similarity-based methods for improving the accuracy of predictions [104].

5.4. Studies of Peptide-Protein Interactions

Databases support studies of peptide-protein interactions through computational methods and experimental validation.

5.4.1. Docking and Molecular Dynamics Simulations

Computational methods, such as docking and molecular dynamics simulations, can be used to predict the interactions between peptides and their target proteins [99,105]. These methods allow researchers to investigate the binding modes and affinities of peptides for their target proteins, providing valuable insights into the molecular mechanisms of peptide action. Chai et al. [13] highlight the use of molecular docking tools to elucidate interactions between bioactive peptides and target proteins, demonstrating the utility of these methods in studying peptide-protein interactions. Hazarika et al. used large-scale docking to predict the binding sites of sORF-encoded peptides on protein surfaces in *Arabidopsis thaliana*, showcasing the application of computational methods for studying peptide-protein interactions [99].

5.4.2. Experimental Validation of Predicted Interactions

Predicted interactions need experimental validation to confirm their biological relevance. Various experimental techniques, such as yeast two-hybrid assays, co-immunoprecipitation, or surface plasmon resonance (SPR), can be used to validate predicted peptide-protein interactions [106]. The experimental validation of predicted interactions is crucial for ensuring the reliability of computational predictions.

6. Future Directions and Challenges in Plant Peptide Database Development

Despite significant progress, several challenges and opportunities remain in the development and utilization of plant peptide databases, challenges such as data standardization, validation and accessibility remain critical for their effective utilization in plant proteomics research [107].

6.1. Data Integration and Standardization

The increasing volume of plant peptide data underscores the need for standardized data formats and integrated omics data. The lack of standardized data formats hinders data exchange and integration across different databases. The development of a universal data standard would greatly improve data interoperability and facilitate the development of more comprehensive and integrated plant peptide databases. Integrating different omics data types, such as genomics, transcriptomics, proteomics, and metabolomics, is crucial for obtaining a more comprehensive understanding of bioactivities of plant peptides [108]. This requires the development of robust data integration methods and the adoption of standardized data formats. The integration of multiple omics datasets can provide valuable insights into the complex interplay between different biological processes [109]. Carroll et al. emphasize the importance of integrating diverse omics information to support proteomic data, highlighting the challenges and benefits of this integrative approach [110].

6.2. Data Quality and Validation

Maintaining data quality and implementing robust validation methods are crucial for the reliability of plant peptide databases [111].

6.2.1. Development of Improved Data Validation Methods

Existing data validation methods can be further refined to enhance data accuracy and reliability. This includes developing more sensitive statistical methods for detecting errors and incorporating more rigorous manual review procedures [112].

6.2.2. Implementation of Quality Control Metrics

The consistent use of quality control metrics across different databases is essential for ensuring data comparability and reliability [113]. This would improve the overall quality and trustworthiness of the data. Tanca et al. highlight the importance of database selection in metaproteomics and suggest strategies for improving the depth and reliability of metaproteomic results. The study emphasizes the need for appropriate filters for taxonomic assignments to minimize false positives [113].

6.2.3. Enhanced Data Accessibility and User Friendliness

Improving data accessibility and user-friendliness is vital for maximizing the impact of plant peptide databases. To maximize the impact of plant peptide databases, developers must prioritize an intuitive User Experience (UX) that bridges the gap between raw bioinformatic data and actionable biological insights. By integrating dynamic data visualization tools and implementing advanced search filters-such as those based on physicochemical properties and functional categories-researchers can pinpoint therapeutic or antimicrobial candidates without navigating complex datasets [113,114].

6.2.4. Development of User-Friendly Web Interfaces

Databases require intuitive and user-friendly web interfaces to facilitate data access and exploration. This includes designing clear navigation menus, incorporating advanced search options, and providing effective data visualization tools [114].

6.2.5. Development of Improved Search and Data Visualization Tools

More sophisticated search and data visualization tools are needed to manage and explore the growing volume of plant peptide data. This includes developing advanced search algorithms that can handle complex queries and implementing more powerful visualization tools that can effectively represent complex datasets [115].

6.3. Addressing Gaps in Coverage

Expanding the coverage of plant peptide databases to include non-model species and diverse peptide types remains a significant challenge. Current databases largely focus on model plant species. Expanding coverage to include non-model plant species is essential for advancing our understanding of plant biology and for harnessing the potential of plant peptides from a broader range of species [116]. Sheehan, D. (2012) emphasizes the importance of next-generation genome sequencing for making non-model organisms more accessible for proteomic studies, highlighting the challenges associated with studying non-model organisms due to the limited availability of genomic information [116]. Current databases may not comprehensively represent the wide diversity of plant peptides. Including a broader range of peptide types, such as cyclic peptides, modified peptides, and peptides with unusual structural features, is needed to capture the full spectrum of plant peptide diversity. To address these challenges, future plant peptide databases could benefit from improved automation in data collection, crowdsourcing of peptide sequences, and the integration of cutting-edge machine learning algorithms to predict novel peptides with high accuracy [117,118]. Additionally, the continued expansion of databases to include less-studied plant species, particularly crops important for global food security, will provide a more comprehensive understanding of the peptide diversity present in the plant kingdom.

7. Conclusion: Plant Peptide Databases as Essential Tools for Plant Biology Research

Plant peptide databases and bioinformatics tools have become essential resources for plant biology research, facilitating the discovery of novel peptides, functional annotation of known peptides, and the study of peptide-protein interactions. Their continued development and refinement are crucial for enhancing our understanding of plant biology and for translating this knowledge into practical applications in agriculture, medicine, and other fields. The challenges discussed above, particularly those concerning data integration, standardization, and expansion of coverage, require sustained attention from the research community. Addressing these challenges will ensure that plant peptide databases continue to serve as powerful tools for advancing plant biology research and further enhance plant peptidomics community to harness the potential of plant peptides for improving human health and sustainable agriculture. Overall, plant peptide databases are poised to remain essential resources, driving innovation and discovery in plant science. They not only provide crucial insights into the molecular mechanisms governing plant biology but also open new avenues for applying this knowledge to improve crop varieties and agricultural practices.

References

- Lindsey, K., Casson, S., & Chilley, P. (2002). Peptides: new signalling molecules in plants. *Trends in plant science*, 7(2), 78–83. [https://doi.org/10.1016/s1360-1385\(01\)02194-x](https://doi.org/10.1016/s1360-1385(01)02194-x)
- Lu, S., & Xiao, F. (2024). Small Peptides: Orchestrators of Plant Growth and Developmental Processes. *International journal of molecular sciences*, 25(14), 7627. <https://doi.org/10.3390/ijms25147627>
- Fedoreyeva, L. I. (2023). Molecular Mechanisms of Regulation of Root Development by Plant Peptides. *Plants*, 12(6), 1320. <https://doi.org/10.3390/plants12061320>
- Katsir, L., Davies, K. A., Bergmann, D. C., & Laux, T. (2011). Peptide signaling in plant development. *Current biology : CB*, 21(9), R356–R364. <https://doi.org/10.1016/j.cub.2011.03.012>
- Ahanger, M. A., Akram, N. A., Ashraf, M., Alyemeni, M. N., Wijaya, L., & Ahmad, P. (2017). Plant responses to environmental stresses-from gene to biotechnology. *AoB PLANTS*, 9(4), plx025. <https://doi.org/10.1093/aobpla/plx025>
- Li, X., Zhu, L., Wang, H., Zhou, X., Wang, M., Li, L., Liu, F., Sun, J., & Xiao, G. (2025). Peptide Hormone-Mediated Regulation of plant development and Environmental Adaptability. *Advanced Science*. <https://doi.org/10.1002/advs.202506590>
- Akbarian, M., Khani, A., Eghbali, S., & Uversky, V. N. (2022). Bioactive Peptides: Synthesis, Sources, Applications, and Proposed Mechanisms of Action. *International journal of molecular sciences*, 23(3), 1445. <https://doi.org/10.3390/ijms23031445>
- Görgüç, A., Gençdağ, E., & Yılmaz, F. M. (2020). Bioactive peptides derived from plant origin by-products: Biological activities and techno-functional utilizations in food developments – A review. *Food Research International*, 136, 109504. <https://doi.org/10.1016/j.foodres.2020.109504>
- Hu, X. L., Lu, H., Hassan, M. M., Zhang, J., Yuan, G., Abraham, P. E., Shrestha, H. K., Villalobos Solis, M. I., Chen, J. G., Tschaplinski, T. J., Doktycz, M. J., Tuskan, G. A., Cheng, Z. M., & Yang, X. (2021). Advances and perspectives in discovery and functional analysis of small secreted proteins in plants. *Horticulture research*, 8(1), 130. <https://doi.org/10.1038/s41438-021-00570-7>
- Silverstein, K. A., Moskal, W. A., Jr, Wu, H. C., Underwood, B. A., Graham, M. A., Town, C. D., & VandenBosch, K. A. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *The Plant journal : for cell and molecular biology*, 51(2), 262–280. <https://doi.org/10.1111/j.1365-313X.2007.03136.x>
- Lease, K. A., & Walker, J. C. (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant physiology*, 142(3), 831–838. <https://doi.org/10.1104/pp.106.086041>
- Sun, Q., Zybaylov, B., Majeran, W., Friso, G., Olinares, P. D., & van Wijk, K. J. (2009). PPDB, the Plant Proteomics Database at Cornell. *Nucleic acids research*, 37(Database issue), D969–D974. <https://doi.org/10.1093/nar/gkn654>
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., & Millar, A. H. (2007). SUBA: the Arabidopsis Subcellular Database. *Nucleic acids research*, 35(Database issue), D213–D218. <https://doi.org/10.1093/nar/gkl863>
- Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W. X., Sun, Q., Van Wijk, K. J., Egelhofer, V., Wienkoop, S., Weckwerth, W., Bruley, C., Rolland, N., Toyoda, T., Nakagami, H., Jones, A. M., Briggs, S. P., Castleden, I., . . . Heazlewood, J. L. (2010). MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *PLANT PHYSIOLOGY*, 155(1), 259–270. <https://doi.org/10.1104/pp.110.168195>
- Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., & Varshney, R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnology Journal*, 18(7), 1507–1525. <https://doi.org/10.1111/pbi.13372>
- Mysore, K. S., Tuori, R. P., & Martin, G. B. (2001). Arabidopsis genome sequence as a tool for functional genomics in tomato. *Genome biology*, 2(1), REVIEWS1003. <https://doi.org/10.1186/gb-2001-2-1-reviews1003>
- Liu, Y., Lu, S., Liu, K., Wang, S., Huang, L., & Guo, L. (2019). Proteomics: a powerful tool to study plant responses to biotic stress. *Plant Methods*, 15(1). <https://doi.org/10.1186/s13007-019-0515-8>
- Araújo, G. S., Ribeiro, G. O., de Souza, S. M. A., Paulo da Silva, G., de Carvalho, G. B. M., Bispo, J. A. C., & Martínez, E. A. (2022). Rice (*Oryza sativa*) Bran and Soybean (*Glycine max*) Meal: Unconventional

- Supplements in the Mead Production. *Food technology and biotechnology*, 60(1), 89–98. <https://doi.org/10.17113/ftb.60.01.22.7183>
19. Baró, A., Saldarelli, P., Saponari, M., Montesinos, E., & Montesinos, L. (2022). *Nicotiana benthamiana* as a model plant host for *Xylella fastidiosa*: Control of infections by transient expression and endotherapy with a bifunctional peptide. *Frontiers in plant science*, 13, 1061463. <https://doi.org/10.3389/fpls.2022.1061463>
 20. Liang, Y., Zhu, W., Chen, S., Qian, J., & Li, L. (2021). Genome-Wide Identification and Characterization of Small Peptides in Maize. *Frontiers in plant science*, 12, 695439. <https://doi.org/10.3389/fpls.2021.695439>
 21. Ruperao, P., Rangan, P., Shah, T., Thakur, V., Kalia, S., Mayes, S., & Rathore, A. (2023). The Progression in Developing Genomic Resources for Crop Improvement. *Life (Basel, Switzerland)*, 13(8), 1668. <https://doi.org/10.3390/life13081668>
 22. Bizzotto, E., Zampieri, G., Treu, L., Filannino, P., Di Cagno, R., & Campanaro, S. (2024). Classification of bioactive peptides: A systematic benchmark of models and encodings. *Computational and Structural Biotechnology Journal*, 23, 2442–2452. <https://doi.org/10.1016/j.csbj.2024.05.040>
 23. Shortreed, M. R., Wenger, C. D., Frey, B. L., Sheynkman, G. M., Scalf, M., Keller, M. P., Attie, A. D., & Smith, L. M. (2015). Global Identification of Protein Post-translational Modifications in a Single-Pass Database Search. *Journal of proteome research*, 14(11), 4714–4720. <https://doi.org/10.1021/acs.jproteome.5b00599>
 24. Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., & van Wijk, K. J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS one*, 3(4), e1994. <https://doi.org/10.1371/journal.pone.0001994>
 25. Mattei, B., Spinelli, F., Pontiggia, D., & De Lorenzo, G. (2016). Comprehensive Analysis of the Membrane Phosphoproteome Regulated by Oligogalacturonides in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.01107>
 26. Heazlewood, J. L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D., & Schulze, W. X. (2008). PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic acids research*, 36(Database issue), D1015–D1021. <https://doi.org/10.1093/nar/gkm812>
 27. Yao, Q., Bollinger, C., Gao, J., Xu, D., & Thelen, J. J. (2012). P3DB: an integrated database for plant protein phosphorylation. *Frontiers in Plant Science*, 3. <https://doi.org/10.3389/fpls.2012.00206>
 28. Xu, X., Zhu, T., Nikonorova, N., & De Smet, I. (2019). Phosphorylation-Mediated Signalling in Plants. 2, 909–932. <https://doi.org/10.1002/9781119312994.APR0702>
 29. Yan, S., Bhawal, R., Yin, Z., Thannhauser, T. W., & Zhang, S. (2022). Recent advances in proteomics and metabolomics in plants. *Molecular Horticulture*, 2(1). <https://doi.org/10.1186/s43897-022-00038-9>
 30. Hammami, R., Ben Hamida, J., Vergoten, G., & Fliss, I. (2009). PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic acids research*, 37(Database issue), D963–D968. <https://doi.org/10.1093/nar/gkn655>
 31. Niarchou A, Alexandridou A, Athanasiadis E, Spyrou G (2013) C-PAmP: Large Scale Analysis and Database Construction Containing High Scoring Computationally Predicted Antimicrobial Peptides for All the Available Plant Species. *PLoS ONE* 8(11): e79728. <https://doi.org/10.1371/journal.pone.0079728>
 32. Yang, Y., Wu, H., Gao, Y., Tong, W., & Li, K. (2023). MFPPDB: a comprehensive multi-functional plant peptide database. *Frontiers in plant science*, 14, 1224394. <https://doi.org/10.3389/fpls.2023.1224394>
 33. Aguilera-Mendoza, L., Marrero-Ponce, Y., Tellez-Ibarra, R., Llorente-Quesada, M. T., Salgado, J., Barigye, S. J., & Liu, J. (2015). Overlap and diversity in antimicrobial peptide databases: compiling a non-redundant set of sequences. *Bioinformatics (Oxford, England)*, 31(15), 2553–2559. <https://doi.org/10.1093/bioinformatics/btv180>
 34. Tang, R., Tan, H., Dai, Y., Li, L., Huang, Y., Yao, H., Cai, Y., & Yu, G. (2023). Application of antimicrobial peptides in plant protection: making use of the overlooked merits. *Frontiers in plant science*, 14, 1139539. <https://doi.org/10.3389/fpls.2023.1139539>
 35. Purohit, K., Reddy, N., & Sunna, A. (2024). Exploring the Potential of Bioactive Peptides: From Natural Sources to Therapeutics. *International journal of molecular sciences*, 25(3), 1391. <https://doi.org/10.3390/ijms25031391>

36. Ahmed, S., Khan, M. S. S., Xue, S., Islam, F., Ikram, A. U., Abdullah, M., Liu, S., Tappiban, P., & Chen, J. (2024). A comprehensive overview of omics-based approaches to enhance biotic and abiotic stress tolerance in sweet potato. *Horticulture Research*, 11(3). <https://doi.org/10.1093/hr/uhae014>
37. Ali, M. H., Mandal, S., Ghorai, M., Lal, M. K., Tiwari, R. K., Kumar, M., Radha, N., Ghosh, A., Al-Tawaha, A. R., Gopalakrishnan, A. V., Shekhawat, M. S., Pandey, D. K., Malik, T., Bursal, E., & Dey, A. (2023). Perspectives of omics and plant microbiome. In *Elsevier eBooks* (pp. 131–144). <https://doi.org/10.1016/b978-0-323-95989-6.00014-0>
38. Riaño-Pachón, D. M., Nagel, A., Neigenfind, J., Wagner, R., Basekow, R., Weber, E., Mueller-Roeber, B., Diehl, S., & Kersten, B. (2009). GabiPD: the GABI primary database—a plant integrative ‘omics’ database. *Nucleic acids research*, 37(Database issue), D954–D959. <https://doi.org/10.1093/nar/gkn611>
39. Liu, L., Liu, E., Hu, Y., Li, S., Zhang, S., Chao, H., Hu, Y., Zhu, Y., Chen, Y., Xie, L., Shen, Y., Wu, L., & Chen, M. (2024). ncPlantDB: a plant ncRNA database with potential ncPEP information and cell type-specific interaction. *Nucleic Acids Research*, 53(D1), D1587–D1594. <https://doi.org/10.1093/nar/gkae1017>
40. Spannagl, M., Nussbaumer, T., Bader, K., Gundlach, H., & Mayer, K. F. (2017). PGSB/MIPS PlantsDB Database Framework for the Integration and Analysis of Plant Genome Data. *Methods in molecular biology (Clifton, N.J.)*, 1533, 33–44. https://doi.org/10.1007/978-1-4939-6658-5_2
41. Gutierrez Reyes, C. D., Alejo-Jacuinde, G., Perez Sanchez, B., Chavez Reyes, J., Onigbinde, S., Mogut, D., Hernández-Jasso, I., Calderón-Vallejo, D., Quintanar, J. L., & Mechref, Y. (2024). Multi Omics Applications in Biological Systems. *Current issues in molecular biology*, 46(6), 5777–5793. <https://doi.org/10.3390/cimb46060345>
42. Das, S., Lecours Boucher, X., Rogers, C., Makowski, C., Chouinard-Decorte, F., Oros Klein, K., Beck, N., Rioux, P., Brown, S. T., Mohaddes, Z., Zweber, C., Foing, V., Forest, M., O’Donnell, K. J., Clark, J., Meaney, M. J., Greenwood, C. M. T., & Evans, A. C. (2018). Integration of “omics” Data and Phenotypic Data Within a Unified Extensible Multimodal Framework. *Frontiers in neuroinformatics*, 12, 91. <https://doi.org/10.3389/fninf.2018.00091>
43. Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W. X., Sun, Q., van Wijk, K. J., Egelhofer, V., Wienkoop, S., Weckwerth, W., Bruley, C., Rolland, N., Toyoda, T., Nakagami, H., Jones, A. M., Briggs, S. P., Castleden, I., Tanz, S. K., ... Heazlewood, J. L. (2011). MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *Plant physiology*, 155(1), 259–270. <https://doi.org/10.1104/pp.110.168195>
44. Mann, G. W., Calley, P. C., Joshi, H. J., & Heazlewood, J. L. (2013). MASCP gator: an overview of the Arabidopsis proteomic aggregation portal. *Frontiers in plant science*, 4, 411. <https://doi.org/10.3389/fpls.2013.00411>
45. Tyagi, A., Pankaj, V., Singh, S., Roy, S., Semwal, M., Shasany, A. K., & Sharma, A. (2019). PlantAFP: a curated database of plant-origin antifungal peptides. *Amino acids*, 51(10-12), 1561–1568. <https://doi.org/10.1007/s00726-019-02792-5>
46. Sheng, P., Xu, M., Zheng, Z., Liu, X., Ma, W., Ding, T., Zhang, C., Chen, M., Zhang, M., Cheng, B., & Zhang, X. (2023). Peptidome and Transcriptome Analysis of Plant Peptides Involved in *Bipolaris maydis* Infection of Maize. *Plants*, 12(6), 1307. <https://doi.org/10.3390/plants12061307>
47. Hellinger, R., Sigurdsson, A., Wu, W., Romanova, E. V., Li, L., Sweedler, J. V., Süßmuth, R. D., & Gruber, C. W. (2023). Peptidomics. *Nature reviews. Methods primers*, 3, 25. <https://doi.org/10.1038/s43586-023-00205-2>
48. Das, D., Jaiswal, M., Khan, F. N., Ahamad, S., & Kumar, S. (2020). PlantPepDB: A manually curated plant peptide database. *Scientific reports*, 10(1), 2194. <https://doi.org/10.1038/s41598-020-59165-2>
49. Willems, P., Horne, A., Van Parys, T., Goormachtig, S., De Smet, I., Botzki, A., Van Breusegem, F., & Gevaert, K. (2019). The Plant PTM Viewer, a central resource for exploring plant protein modifications. *The Plant Journal*, 99(4), 752–762. <https://doi.org/10.1111/tpj.14345>
50. MacFarlane, A., Russell-Rose, T., & Shokraneh, F. (2022). Search strategy formulation for systematic reviews: Issues, challenges and opportunities. *Intelligent Systems With Applications*, 15, 200091. <https://doi.org/10.1016/j.iswa.2022.200091>

51. Xue, H., Zhang, Q., Wang, P., Cao, B., Jia, C., Cheng, B., Shi, Y., Guo, W., Wang, Z., Liu, Z., & Cheng, H. (2021). qPTMplants: an integrative database of quantitative post-translational modifications in plants. *Nucleic Acids Research*, 50(D1), D1491–D1499. <https://doi.org/10.1093/nar/gkab945>
52. Rashid, M., Omar, M., & Mohanta, T. K. (2023). FungiProteomeDB: a database for the molecular weight and isoelectric points of the fungal proteomes. *Database*, 2023. <https://doi.org/10.1093/database/baad004>
53. Schuler, G. D., Epstein, J. A., Ohkawa, H., & Kans, J. A. (1996). Entrez: molecular biology database and retrieval system. *Methods in enzymology*, 266, 141–162. [https://doi.org/10.1016/s0076-6879\(96\)66012-1](https://doi.org/10.1016/s0076-6879(96)66012-1)
54. Kitts, P. A., Church, D. M., Thibaud-Nissen, F., Choi, J., Hem, V., Sapojnikov, V., Smith, R. G., Tatusova, T., Xiang, C., Zherikov, A., DiCuccio, M., Murphy, T. D., Pruitt, K. D., & Kimchi, A. (2016). Assembly: a resource for assembled genomes at NCBI. *Nucleic acids research*, 44(D1), D73–D80. <https://doi.org/10.1093/nar/gkv1226>
55. Jeanquartier, F., Jean-Quartier, C., & Holzinger, A. (2015). Integrated web visualizations for protein-protein interaction databases. *BMC bioinformatics*, 16(1), 195. <https://doi.org/10.1186/s12859-015-0615-z>
56. Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. (2023). A comprehensive review of data visualization tools: features, strengths, and weaknesses. *International Journal of Computer Engineering in Research Trends*, 10(1), 10–20. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0102>
57. Zhang, K., Teng, D., Mao, R., Yang, N., Hao, Y., & Wang, J. (2023). Thinking on the Construction of Antimicrobial Peptide Databases: Powerful Tools for the Molecular Design and Screening. *International journal of molecular sciences*, 24(4), 3134. <https://doi.org/10.3390/ijms24043134>
58. Shi, G., Kang, X., Dong, F., Liu, Y., Zhu, N., Hu, Y., Xu, H., Lao, X., & Zheng, H. (2022). DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic acids research*, 50(D1), D488–D496. <https://doi.org/10.1093/nar/gkab651>
59. Qin, D., Bo, W., Zheng, X., Hao, Y., Li, B., Zheng, J., & Liang, G. (2022). DFBP: a comprehensive database of food-derived bioactive peptides for peptidomics research. *Bioinformatics*, 38(12), 3275–3280. <https://doi.org/10.1093/bioinformatics/btac323>
60. Xu, Y. Y., Zhou, H., Murphy, R. F., & Shen, H. B. (2021). Consistency and variation of protein subcellular location annotations. *Proteins*, 89(2), 242–250. <https://doi.org/10.1002/prot.26010>
61. Liu, J., Yang, M., Yu, Y., Xu, H., Wang, T., Li, K., & Zhou, X. (2025). Advancing bioinformatics with large language models: components, applications and perspectives. *ArXiv*, arXiv:2401.04155v2.
62. Kress, A., Poch, O., Lecompte, O., & Thompson, J. D. (2023). Real or fake? Measuring the impact of protein annotation errors on estimates of domain gain and loss events. *Frontiers in bioinformatics*, 3, 1178926. <https://doi.org/10.3389/fbinf.2023.1178926>
63. Nguyen, T., Nguyen, H., & Nguyen-Hoang, T. (2025). Data quality management in Big Data: Strategies, tools, and educational implications. *Journal of Parallel and Distributed Computing*, 105067. <https://doi.org/10.1016/j.jpdc.2025.105067>
64. Gliklich, R. E., Leavy, M. B., & Dreyer, N. A. (2020, September 1). *Obtaining data and quality assurance. Registries for Evaluating Patient Outcomes: A User's Guide - NCBI Bookshelf*. https://www.ncbi.nlm.nih.gov/books/NBK562556/#_ncbi_dlg_citbx_NBK562556
65. Sinha, A., Hripcsak, G., & Markatou, M. (2009). Large datasets in biomedicine: a discussion of salient analytic issues. *Journal of the American Medical Informatics Association : JAMIA*, 16(6), 759–767. <https://doi.org/10.1197/jamia.M2780>
66. Martin-Sanchez, F., & Verspoor, K. (2014). Big data in medicine is driving big changes. *Yearbook of medical informatics*, 9(1), 14–20. <https://doi.org/10.15265/IY-2014-0020>
67. Quiroz, C., Saavedra, Y. B., Armijo-Galdames, B., Amado-Hinojosa, J., Olivera-Nappa, Á., Sanchez-Daza, A., & Medina-Ortiz, D. (2021). Peptipedia: a user-friendly web application and a comprehensive database for peptide research supported by Machine Learning approach. *Database : the journal of biological databases and curation*, 2021, baab055. <https://doi.org/10.1093/database/baab055>
68. Blair-Early, A., & Zender, M. (2008). User interface design principles for interaction design. *Design Issues*, 24(3), 85–107. <https://doi.org/10.1162/desi.2008.24.3.85>
69. Kimball, M. A. (2013). Visual Design Principles: An Empirical Study of Design lore. *Journal of Technical Writing and Communication*, 43(1), 3–41. <https://doi.org/10.2190/tw.43.1.b>

70. Zhang, H. (2016). Overview of sequence data formats. *Methods in Molecular Biology*, 3–17. https://doi.org/10.1007/978-1-4939-3578-9_1
71. Parnas, D. L. (2010). Precise Documentation: The Key to Better Software. In *Springer eBooks* (pp. 125–148). https://doi.org/10.1007/978-3-642-15187-3_8
72. Lease, K. A., & Walker, J. C. (2010). Bioinformatic identification of plant peptides. *Methods in molecular biology (Clifton, N.J.)*, 615, 375–383. https://doi.org/10.1007/978-1-60761-535-4_26
73. Birhanu A. G. (2023). Mass spectrometry-based proteomics as an emerging tool in clinical laboratories. *Clinical proteomics*, 20(1), 32. <https://doi.org/10.1186/s12014-023-09424-x>
74. Abdul-Khalek, N., Wimmer, R., Overgaard, M. T., & Gregersen Echers, S. (2023). Insight on physicochemical properties governing peptide MS1 response in HPLC-ESI-MS/MS: A deep learning approach. *Computational and structural biotechnology journal*, 21, 3715–3727. <https://doi.org/10.1016/j.csbj.2023.07.027>
75. Héthelyi, E., Tétényi, P., Dabi, E., & Dános, B. (1987). The role of mass spectrometry in medicinal plant research. *Biomedical & environmental mass spectrometry*, 14(11), 627–632. <https://doi.org/10.1002/bms.1200141110>
76. Moini M. (2002). Capillary electrophoresis mass spectrometry and its application to the analysis of biological mixtures. *Analytical and bioanalytical chemistry*, 373(6), 466–480. <https://doi.org/10.1007/s00216-002-1283-1>
77. Moini M. (2004). Capillary electrophoresis-electrospray ionization mass spectrometry of amino acids, peptides, and proteins. *Methods in molecular biology (Clifton, N.J.)*, 276, 253–290. <https://doi.org/10.1385/1-59259-798-X:253>
78. Mant, C. T., Chen, Y., Yan, Z., Popa, T. V., Kovacs, J. M., Mills, J. B., Tripet, B. P., & Hodges, R. S. (2007). HPLC analysis and purification of peptides. *Methods in molecular biology (Clifton, N.J.)*, 386, 3–55. https://doi.org/10.1007/978-1-59745-430-8_1
79. Shaw, C. (2003). Reverse-Phase HPLC Purification of Peptides from Natural Sources for Structural Analysis. *Humana Press eBooks*, 101–108. <https://doi.org/10.1385/0-89603-353-8:101>
80. Hong, P., Koza, S., & Bouvier, E. S. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of liquid chromatography & related technologies*, 35(20), 2923–2950. <https://doi.org/10.1080/10826076.2012.743724>
81. Hayes, R., Ahmed, A., Edge, T., & Zhang, H. (2014). Core-shell particles: Preparation, fundamentals and applications in high performance liquid chromatography. *Journal of Chromatography A*, 1357, 36–52. <https://doi.org/10.1016/j.chroma.2014.05.010>
82. Scriba G. K. (2016). Separation of Peptides by Capillary Electrophoresis. *Methods in molecular biology (Clifton, N.J.)*, 1483, 365–391. https://doi.org/10.1007/978-1-4939-6403-1_18
83. Misra, B. B., & Van Der Hoof, J. J. J. (2015). Updates in metabolomics tools and resources: 2014–2015. *Electrophoresis*, 37(1), 86–110. <https://doi.org/10.1002/elps.201500417>
84. Chen, C., Hou, J., Tanner, J. J., & Cheng, J. (2020). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *International journal of molecular sciences*, 21(8), 2873. <https://doi.org/10.3390/ijms21082873>
85. Ferreira, R., Amado, F., & Vitorino, R. (2023). Empowering peptidomics: utilizing computational tools and approaches. *Bioanalysis*, 15(21), 1315–1325. <https://doi.org/10.4155/bio-2023-0102>
86. Yang, J., & Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Current protocols in bioinformatics*, 52, 5.8.1–5.8.15. <https://doi.org/10.1002/0471250953.bi0508s52>
87. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., ... Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
88. Zhang, Y., & Sanner, M. F. (2019). Docking Flexible Cyclic Peptides with *AutoDock CrankPep*. *Journal of chemical theory and computation*, 15(10), 5161–5168. <https://doi.org/10.1021/acs.jctc.9b00557>

89. Tubert-Brohman, I., Sherman, W., Repasky, M., & Beuming, T. (2013). Improved docking of polypeptides with Glide. *Journal of chemical information and modeling*, 53(7), 1689–1699. <https://doi.org/10.1021/ci400128m>
90. Minkiewicz, P., Iwaniak, A., & Darewicz, M. (2019). BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. *International journal of molecular sciences*, 20(23), 5978. <https://doi.org/10.3390/ijms20235978>
91. Misra B. B. (2018). Updates on resources, software tools, and databases for plant proteomics in 2016–2017. *Electrophoresis*, 39(13), 1543–1557. <https://doi.org/10.1002/elps.201700401>
92. Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS medicine*, 2(10), e267. <https://doi.org/10.1371/journal.pmed.0020267>
93. Bogdanow, B., Zauber, H., & Selbach, M. (2016). Systematic Errors in Peptide and Protein Identification and Quantification by Modified Peptides. *Molecular & cellular proteomics : MCP*, 15(8), 2791–2801. <https://doi.org/10.1074/mcp.M115.055103>
94. Caliskan, A., Dangwal, S., & Dandekar, T. (2023). Metadata integrity in bioinformatics: Bridging the gap between data and knowledge. *Computational and structural biotechnology journal*, 21, 4895–4913. <https://doi.org/10.1016/j.csbj.2023.10.006>
95. Johnson, R. S., Searle, B. C., Nunn, B. L., Gilmore, J. M., Phillips, M., Amemiya, C. T., Heck, M., & MacCoss, M. J. (2020). Assessing Protein Sequence Database Suitability Using *De Novo* Sequencing. *Molecular & cellular proteomics : MCP*, 19(1), 198–208. <https://doi.org/10.1074/mcp.TIR119.001752>
96. Kinsinger, C. R., Apffel, J., Baker, M., Bian, X., Borchers, C. H., Bradshaw, R., Brusniak, M. Y., Chan, D. W., Deutsch, E. W., Domon, B., Gorman, J., Grimm, R., Hancock, W., Hermjakob, H., Horn, D., Hunter, C., Kolar, P., Kraus, H. J., Langen, H., Linding, R., ... Rodriguez, H. (2012). Recommendations for mass spectrometry data quality metrics for open access data (corollary to the Amsterdam Principles). *Journal of proteome research*, 11(2), 1412–1419. <https://doi.org/10.1021/pr201071t>
97. Lease, K. A., & Walker, J. C. (2009). Bioinformatic identification of plant peptides. *Methods in Molecular Biology*, 375–383. https://doi.org/10.1007/978-1-60761-535-4_26
98. Hellinger, R., Koehbach, J., Soltis, D. E., Carpenter, E. J., Wong, G. K., & Gruber, C. W. (2015). Peptidomics of Circular Cysteine-Rich Plant Peptides: Analysis of the Diversity of Cyclotides from *Viola tricolor* by Transcriptome and Proteome Mining. *Journal of proteome research*, 14(11), 4851–4862. <https://doi.org/10.1021/acs.jproteome.5b00681>
99. Hazarika, R. R., Sostaric, N., Sun, Y., & Van Noort, V. (2018). Large-scale docking predicts that sORF-encoded peptides may function through protein-peptide interactions in *Arabidopsis thaliana*. *PLoS ONE*, 13(10), e0205179. <https://doi.org/10.1371/journal.pone.0205179>
100. Porto, W. F., Souza, V. A., Nolasco, D. O., & Franco, O. L. (2012). In silico identification of novel hevein-like peptide precursors. *Peptides*, 38(1), 127–136. <https://doi.org/10.1016/j.peptides.2012.07.025>
101. Chamoli, T., Khera, A., Sharma, A., Gupta, A., Garg, S., Mamgain, K., Bansal, A., Verma, S., Gupta, A., Alajangi, H. K., Singh, G., & Barnwal, R. P. (2022). Peptide Utility (PU) search server: A new tool for peptide sequence search from multiple databases. *Heliyon*, 8(12), e12283. <https://doi.org/10.1016/j.heliyon.2022.e12283>
102. Kourelis, J., Kaschani, F., Grosse-Holz, F. M., Homma, F., Kaiser, M., & Van Der Hoorn, R. a. L. (2019). A homology-guided, genome-based proteome for improved proteomics in the allopolyploid *Nicotiana benthamiana*. *BMC Genomics*, 20(1). <https://doi.org/10.1186/s12864-019-6058-6>
103. Strabala, T. J., Phillips, L., West, M., & Stanbra, L. (2014). Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. *BMC Plant Biology*, 14(1), 47. <https://doi.org/10.1186/1471-2229-14-47>
104. Kaur, D., Arora, A., Vigneshwar, P., & Raghava, G. P. (2023). Prediction of peptide hormones using an ensemble of machine learning and similarity-based methods. *bioRxiv (Cold Spring Harbor Laboratory)*. <https://doi.org/10.1101/2023.05.15.540764>
105. Chai, T., Ee, K., Kumar, D. T., Manan, F. A., & Wong, F. (2020). Plant Bioactive peptides: Current status and Prospects towards use on human health. *Protein and Peptide Letters*, 28(6), 623–642. <https://doi.org/10.2174/0929866527999201211195936>

106. Brückner, A., Polge, C., Lentze, N., Auerbach, D., & Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *International journal of molecular sciences*, 10(6), 2763–2788. <https://doi.org/10.3390/ijms10062763>
107. Chandramouli, K., & Qian, P. Y. (2009). Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics : HGP, 2009*, 239204. <https://doi.org/10.4061/2009/239204>
108. Chervitz, S. A., Deutsch, E. W., Field, D., Parkinson, H., Quackenbush, J., Rocca-Serra, P., Sansone, S. A., Stoeckert, C. J., Jr, Taylor, C. F., Taylor, R., & Ball, C. A. (2011). Data standards for Omics data: the basis of data sharing and reuse. *Methods in molecular biology (Clifton, N.J.)*, 719, 31–69. https://doi.org/10.1007/978-1-61779-027-0_2
109. Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and biology insights*, 14, 1177932219899051. <https://doi.org/10.1177/1177932219899051>
110. Carroll, A. W., Joshi, H. J., & Heazlewood, J. L. (2013). Managing the green proteomes for the next decade of plant research. *Frontiers in Plant Science*, 4. <https://doi.org/10.3389/fpls.2013.00501>
111. Rappsilber, J., Bruce, J., Combe, C., Fried, S. D., Graziadei, A., Heck, A. J. R., Iacobucci, C., Leitner, A., Mechtler, K., Novak, P., O'Reilly, F., Schriemer, D. C., Sinz, A., Stengel, F., & Thalassinou, K. (2025). A Roadmap for Improving Reliability and Data Sharing in Crosslinking Mass Spectrometry. *Molecular & cellular proteomics : MCP*, 24(8), 101024. Advance online publication. <https://doi.org/10.1016/j.mcpro.2025.101024>
112. Wang G. (2015). Improved methods for classification, prediction, and design of antimicrobial peptides. *Methods in molecular biology (Clifton, N.J.)*, 1268, 43–66. https://doi.org/10.1007/978-1-4939-2285-7_3
113. Tanca, A., Palomba, A., Deligios, M., Cubeddu, T., Fraumene, C., Biosa, G., Pagnozzi, D., Addis, M. F., & Uzzau, S. (2013). Evaluating the Impact of Different Sequence Databases on Metaproteome Analysis: Insights from a Lab-Assembled Microbial Mixture. *PLoS ONE*, 8(12), e82981. <https://doi.org/10.1371/journal.pone.0082981>
114. Aplakidou, E., Vergoulidis, N., Chasapi, M., Venetsianou, N. K., Kokoli, M., Panagiotopoulou, E., Iliopoulos, I., Karatzas, E., Pafilis, E., Georgakopoulos-Soares, I., Kyrpides, N. C., Pavlopoulos, G. A., & Baltoumas, F. A. (2024). Visualizing metagenomic and metatranscriptomic data: A comprehensive review. *Computational and Structural Biotechnology Journal*, 23, 2011–2033. <https://doi.org/10.1016/j.csbj.2024.04.060>
115. Morabito, A., De Simone, G., Pastorelli, R., Brunelli, L., & Ferrario, M. (2025). Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review. *Journal of Translational Medicine*, 23(1). <https://doi.org/10.1186/s12967-025-06446-x>
116. Sheehan, D. (2012). Next-Generation genome sequencing makes Non-Model organisms increasingly accessible for proteomic studies: Some implications for ecotoxicology. *Journal of Proteomics & Bioinformatics*, 06(01). <https://doi.org/10.4172/jpb.10000e21>
117. Cao, Y., Yang, P., & Li, M. (2025). Research progress of peptides discovery and function in resistance to abiotic stress in plant. *Stress Biology*, 5(1). <https://doi.org/10.1007/s44154-025-00220-1>
118. Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., & Cammue, B. P. (2015). The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. *The Plant cell*, 27(8), 2095–2118. <https://doi.org/10.1105/tpc.15.00440>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.