Journal of Agriculture and Food Research 30 (2026) 103162

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

# QSAR modeling and machine learning for antifungal peptide discovery in crop protection: Translational challenges and future perspectives

Riccardo Zanni [a,b,*], Maria Galvez-Llompart [a], Ignacio Bueso-Bordils [c], Alejandro Perez-Garcia [d], Dolores Fernandez-Ortuño [d], Facundo Perez-Giménez [b]

[a] Department of Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine, Faculty of Pharmacy, University of Valencia, Spain
[b] Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, Spain
[c] Department of Physical Chemistry, University Cardenal Herrera CEU, Spain
[d] Department of Microbiology, University of Málaga, Institute of Subtropical and Mediterranean Horticulture "La Mayora" (IHSM-UMA-CSIC), Málaga, Spain

ARTICLE INFO

*Keywords:*
Antifungal peptides
QSAR
Machine learning
Protein language models
Biopesticides
Crop protection

ABSTRACT

Fungicide resistance and restrictions on conventional agrochemicals are increasing the need for sustainable crop-protection strategies. Antifungal peptides (AFPs) are considered promising candidates due to their multiple mechanisms of action, biodegradability, and suitability for environmentally sustainable disease control. However, their application in agriculture remains limited by poor field stability, suboptimal delivery systems, high production costs, and the lack of predictive models validated against phytopathogenic fungi. This review collected and critically compared QSAR and machine learning (ML) studies published between 2015 and 2025, focusing on their modeling strategies, predictive performance, limitations, and translational requirements for antifungal peptide discovery in crop protection. The available literature remains dominated by *Candida* spp. and broadly biomedical antifungal datasets, highlighting the translational gap between biomedical AFP prediction and agricultural application. Eight representative studies were compared in terms of dataset size, endpoint, descriptors, algorithms, target fungi, predictive performance, and validation approaches. Dataset size ranged from 45 to 60 peptides in small congeneric series to large-scale collections containing up to 17,514 sequences. Classification models generally reported validation or test accuracies of 0.86–0.96 and AUC values of 0.91–0.99, whereas regression models showed more variable external or test performance, with $R^2$ values of approximately 0.50–0.87. The reviewed studies illustrate a diverse methodological landscape, from interpretable QSAR models to ML classifiers and deep-learning pipelines based on protein-language-model sequence representation. Overall, the principal barrier to agricultural translation is not algorithm availability, but the lack of curated plant-pathogen-specific datasets, standardized agricultural endpoints, and external benchmarks incorporating field-relevant traits such as protease resistance, UV stability, leaf-surface persistence, formulation compatibility, phytotoxicity, and effects on beneficial microorganisms.

## 1. Introduction

### 1.1. Fungal diseases and the need for alternative crop-protection strategies

Food security, as defined at the 2009 World Summit on Food Security, exists "when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active and healthy life." Achieving this security, however, depends on a fragile foundation: a limited number of major crop species that are persistently threatened by a spectrum of phytopathogenic fungi. The agronomic and economic burden of these diseases is staggering [1].

Recent analyses estimate that fungal pathogens destroy up to 23% of global crop yields annually, a massive pre- and post-harvest loss that significantly hinders the transformation of agricultural systems needed to meet future food demand [2,3]. This challenge is monumental, as the global population is projected to exceed 10 billion between 2080 and 2100, driving continuously growing demand for food [4].

* Corresponding author. Department of Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine, Faculty of Pharmacy, University of Valencia, Spain.
*E-mail address:* riccardo.zanni@uv.es (R. Zanni).

https://doi.org/10.1016/j.jafr.2026.103162
Received 18 May 2026; Received in revised form 2 July 2026; Accepted 21 July 2026
Available online 24 July 2026

Fungal pathogens are a primary driver of these yield declines, responsible for an estimated 70–80% of all crop diseases [5], with an annual economic toll exceeding $200 billion USD [6]. The severity of this burden is crystallized in the devastation caused by specific diseases: wheat rusts and Fusarium head blight continually threaten staple cereal production [7], while powdery mildew diseases devastate a vast range of high-value crops, from grapes to cucurbits. Pathogens such as *Fusarium* spp., *Botrytis cinerea, Magnaporthe oryzae, Alternaria* spp., continue to challenge disease management [8,9], while resistance to conventional fungicides like sterol demethylation inhibitors (DMIs) and succinate dehydrogenase inhibitors (SDHIs), is now widespread, with over 50% of US and EU field isolates showing reduced sensitivity [10,11].

This crisis is intensified by stringent new regulations that restrict the use of many existing agrochemicals, such as EU restrictions on triazoles under Regulation EC 1107/2009, alongside public backlash over environmental toxicity [12]. Consequently, farmers are left with increasingly few reliable tools.

These compounding challenges underscore the urgent need for innovative alternative solutions, such as natural or synthetic peptide-based compounds, which offer the potential to combine effective disease control with environmental compatibility.

### *1.2. Antifungal peptides (AFPs) as a sustainable solution for crop protection*

AFPs are attractive candidates for sustainable crop protection because they combine target selectivity, biodegradability, and mechanisms of action that differ from many conventional small-molecule fungicides (Fig. 1) [13–19]. Recent studies have highlighted the relevance of plant-derived antimicrobial peptides, including their mechanisms of action, safety considerations, and potential integration into crop-protection strategies [20].

The activity of antifungal peptides can involve membrane disruption, interference with fungal cell-wall components, intracellular targeting, oxidative stress induction, and activation of plant defense pathways [16, 21–28]. These mechanisms are relevant to phytopathogenic fungi because they intersect with conserved fungal structures and processes, including membranes enriched in sterols and cell envelopes containing chitin, β-glucans, and mannoproteins. Representative mechanisms and experimentally validated examples are summarized in Table 1.

This mechanistic diversity provides a biological basis for the rational design of peptide-based fungicides and supports their potential contribution to resistance-management strategies. However, the same physicochemical properties that make AFPs environmentally attractive also create a formulation and delivery trade-off. Environmental biodegradability is desirable after application because it may reduce long-term residue accumulation in soil, water, and harvested products; however, premature degradation before the peptide reaches or persists at the fungal target can reduce efficacy and require higher or more frequent application rates. Therefore, agricultural AFP design must balance two competing requirements: sufficient stability to remain active on plant surfaces, in the rhizosphere, or during post-harvest storage, and sufficient degradability to avoid persistent environmental residues. Furthermore, in crop-protection settings, peptide performance can be affected by UV radiation, plant and microbial proteases, fluctuating humidity and pH, adsorption to waxy cuticles, and formulation-dependent delivery [15,29,30]. Therefore, agricultural AFP discovery cannot rely only on intrinsic antifungal potency; it must also consider environmental stability, persistence, delivery, phytotoxicity, and selectivity toward beneficial plant-associated microorganisms [15,17,18].

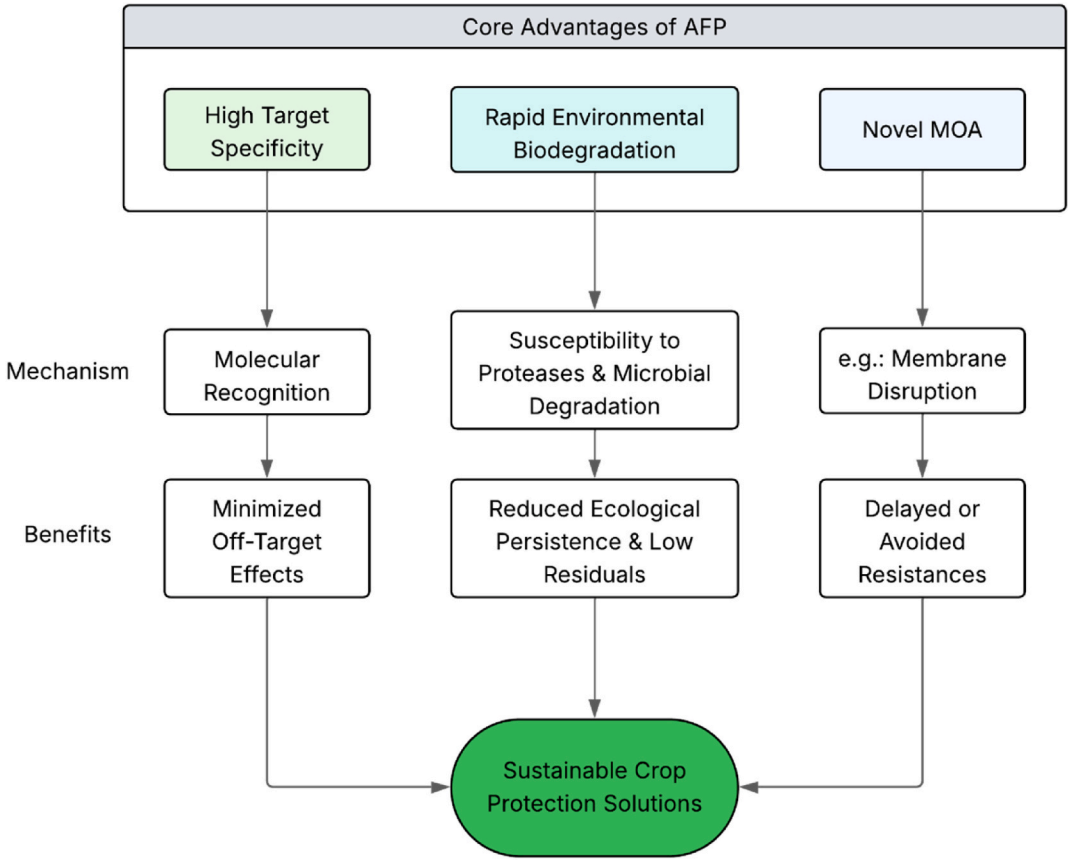


**Fig. 1.** Core advantages of employing antifungal peptides (AFPs) in sustainable crop protection strategies.

**Table 1**
Experimentally validated antifungal mechanisms supporting the rational design of fungicidal peptides for sustainable crop protection.

| Mechanism of Action | Description | Representative Peptides | Confirmed Targets/Pathogens | Key References |
|---|---|---|---|---|
| **Direct antifungal activity** | | | | |
| ***Membrane Disruption*** | Peptides bind to fungal membranes, form pores or cause lysis | Plant defensins, thionins, hevein-like peptides | *Botrytis cinerea*, *Fusarium* spp. | [22–25] |
| ***Cell Wall Biosynthesis Inhibition*** | Interference with chitin synthase or β-glucan synthase | Synthetic chitin-binding peptides | *Colletotrichum*, *Alternaria* | [24,26] |
| ***Intracellular Targeting*** | Peptides penetrate cells and inhibit DNA/RNA or protein synthesis | Nuclear-localizing synthetic peptides | *Magnaporthe oryzae*, *Botrytis cinerea* | [27,28] |
| ***Oxidative Stress Induction*** | Triggers ROS accumulation leading to cellular damage | Microbial-derived AFPs | *Fusarium*, *Rhizoctonia solani* | [26] |
| **Indirect antifungal activity** | | | | |
| ***Host Immunomodulation*** | Activates plant defense pathways | Systemin, defensins | Broad-spectrum fungal resistance | [27] |

### *1.3. QSAR and machine learning applications in antifungal peptide development*

Antifungal peptide discovery uses complementary experimental and computational strategies, including empirical peptide-library screening, phage display, structural biology, molecular docking, and molecular dynamics simulations. These methods address different stages of the discovery pipeline. Empirical screening and phage display can identify active or target-binding peptides, whereas structural and simulation-based approaches can provide detailed mechanistic insight into peptide conformation, membrane interaction, target engagement, and candidate refinement. However, these approaches are experimentally or computationally intensive and are not always suited to first-pass prioritization across large peptide sequence spaces. The present review focuses specifically on QSAR and ML workflows because they provide a scalable framework for integrating peptide sequences, molecular descriptors, activity annotations, and validation metrics to identify structure–activity patterns, screen virtual peptide libraries, and prioritize candidates for subsequent experimental, structural, and formulation studies [30–34].

Building on the translational barriers introduced above, QSAR and ML approaches are particularly relevant for agricultural AFP discovery because they can support candidate prioritization beyond intrinsic antifungal potency. In principle, predictive models may be used to identify motifs associated with proteolytic susceptibility, prioritize peptides with improved physicochemical profiles, and guide the selection of candidates for synthesis and validation [29–31]. Nevertheless, these applications require appropriate datasets and endpoints.

In addition, production cost remains a critical barrier. The chemical synthesis or recombinant expression of peptides is still generally more expensive than the mass production of conventional small-molecule fungicides [35]. ML-driven optimization and computational design may help streamline the development pipeline, from *de novo* peptide design to the prediction of candidate sequences with improved synthetic accessibility, stability, or production feasibility [36,37]. For instance, the integration of deep learning with cell-free protein synthesis has been proposed as a time-, cost-, and labor-effective platform for rapid peptide production and functional screening [38]. Similarly, computational models may support the optimization of microbial or recombinant production systems for naturally derived antifungal agents, including lipopeptides and polyketides [21].

QSAR and ML workflows can also support delivery and formulation development by relating peptide physicochemical properties to carrier compatibility, solubility, stability, and release behavior. Computational models can help prioritize candidates for encapsulation, adjuvant combination, or formulation strategies intended to improve leaf adhesion, cuticle interaction, and duration of activity [15,30]. This view is consistent with recent analyses of enzymes, proteins, and peptides as emerging biosolutions for crop production, which emphasize both their potential and the practical challenges associated with stability, formulation, delivery, and field deployment [39]. Thus, QSAR and ML should be treated as a prioritization framework rather than a replacement for experimental validation, structural analysis, or formulation studies.

The aim of this review is not to introduce a new QSAR and ML strategy or to report a newly trained predictive model. Instead, it provides a critical synthesis of current QSAR and ML approaches applied to antifungal peptide discovery during the last ten years, with particular attention to their relevance for agricultural and crop-protection applications. Specifically, the review compares published models in terms of datasets, endpoints, descriptors, algorithms, validation strategies, and target fungal species; identifies the current predominance of biomedical datasets and *Candida* spp. centered validation, discussing the additional requirements needed for agricultural translation, including plant-pathogen-specific benchmarking, environmental stability, delivery, phytotoxicity, and formulation compatibility. In this sense, the contribution of the review lies in defining the translational gap and outlining a structured roadmap for adapting existing computational peptide-design strategies to plant fungal disease control.

### *1.4. Methodological framework for the literature review: search strategy and study selection*

This manuscript was designed as a critical narrative review with a structured comparative analysis of QSAR and machine learning studies relevant to antifungal peptide discovery. The literature search was performed using PubMed, Scopus, and SciFinder Scholar as the primary bibliographic databases. ResearchRabbit was used as a supplementary citation-mapping tool to identify additional relevant studies and citation networks. The last search update was conducted on May 17, 2026.

The search targeted studies published between 2015 and 2025 addressing QSAR, machine learning, deep learning, molecular descriptors, or computational peptide-design strategies applied to antifungal or antimicrobial peptide discovery. Search queries were constructed using Boolean operators as follows: (“antifungal peptide” OR “antimicrobial peptide” OR “fungicide” OR “antifungal” OR “peptide”) AND (“QSAR” OR “quantitative structure-activity relationship” OR “machine learning” OR “deep learning” OR “molecular descriptors” OR “peptide design”).

Studies were included if they met the following criteria: (i) development or application of QSAR, machine learning, or deep-learning models for antifungal peptide prediction or design; (ii) use of peptide descriptors, sequence-derived features, physicochemical and structural descriptors, or protein-language-model embeddings for classification or regression tasks involving antifungal and antimicrobial peptides; (iii) reporting of performance metrics such as accuracy, AUC, MCC, F1-score, $R^2$, $Q^2$, $R^2$pred, or external/test-set validation; or (iv) relevance to the translational potential of peptide-based antifungal strategies for crop protection.

## 2. QSAR and machine learning: A decade accelerating the next generation of antifungal peptides

The integration of quantitative structure-activity relationship (QSAR) modeling with machine learning (ML) supports peptide-based discovery by enabling scalable candidate prioritization before synthesis and experimental testing [40,41]. QSAR modeling strategies can accelerate antifungal peptide discovery by establishing a quantitative relationship between peptide structural features and antifungal activity, thereby supporting the rational design and optimization of candidates with potential fungicidal properties. Platforms such as PepQSAR, for example, have demonstrated the ability to screen thousands of peptide sequences *in silico*, facilitating the identification of candidates with desired biological properties [31]. Although the screening of libraries exceeding $10^6$ compounds remains challenging, continuous improvements in ML computational resources are progressively increasing the scale and efficiency of peptide prioritization workflows. These advances highlight the growing role of ML as a predictive tool for accelerating AFP discovery [42,43].

Beyond conventional QSAR approaches, the incorporation of deep-learning and generative models has further expanded the possibilities of computational peptide design by enabling the exploration of novel sequence spaces beyond naturally occurring peptides. Generative ML models can create *de novo* peptide candidates with tailored properties, significantly expanding the chemical space beyond natural sequences and accelerating discovery workflows [44,45]. Recent studies illustrate this emerging trend. Li et al. (2024) developed the AMPd-Up platform, which employed a recurrent neural network (RNN) to generate novel peptide sequences, achieving a high rate of experimental confirmation against bacterial targets [46]. Similarly, Wang et al. (2024) introduced ProT-Diff, which combines a pretrained protein language model with a diffusion model to generate candidate AMPs and AFPs, showing high synthesis success and therapeutic efficacy in a mouse model [47]. Zervou et al. (2024) used feedback generative adversarial networks (FBGAN) to design peptides. Their model incorporates a feedback loop from classifiers like ESM2 (Evolutionary Scale Modeling), which improved accuracy and efficiency in generating promising antimicrobial candidates [48].

One of the most recent examples illustrating the application of these computational advances to antifungal peptide discovery is the Fung-AI pipeline developed by Berman et al., this approach integrates generative adversarial network (GAN)-based sequence generation with *in silico* classifiers for the *de novo* design and prioritization of antifungal peptides. From approximately 10,000 generated sequences, selected candidates exhibited mild *in vitro* activity against *Fusarium graminearum* (a wheat pathogen of relevance to the agri-food industry) and *Candida albicans*. This study provides a proof-of-principle for AI-driven antifungal peptide discovery by combining sequence generation, computational filtering, experimental validation, and preliminary toxicity assessment [49].

Collectively, these studies highlight the integration of QSAR and ML approaches within a broader computational framework for antifungal peptide discovery, as conceptually summarized in Fig. 2. This workflow outlines the interconnected stages commonly applied in computational AFP discovery, ranging from sequence representation and model development to candidate prioritization and experimental validation. In addition, QSAR and ML approaches for AFP discovery should adhere to established best-practice guidelines, where model robustness, external validation, applicability domain assessment, and transparent reporting are essential requirements to ensure that computational predictions can reliably support experimental prioritization [50].

As shown in Fig. 2, four key stages can be identified in the discovery of antifungal peptides using QSAR and ML methodologies: (1) peptide dataset collection; (2) molecular descriptor calculation; (3) model development and validation; and (4) virtual screening and selection of novel candidate peptides [33,50]. Among these, the initial and most critical step is dataset construction (1), which involves the collection, curation, and standardization of raw peptide data. Indeed, the predictive performance of the entire workflow largely depends on the quality, consistency, and representativeness of the training dataset, which may include quantitative and/or qualitative annotations of antifungal activity derived from peptide databases. For classification tasks, dataset assembly should not rely exclusively on repositories enriched in known active antimicrobial or antifungal peptides. Instead, active sequences from databases such as DRAMP, APD3, CAMP, DBAASP, or StarPepDB should be complemented by carefully selected inactive, weakly active, or decoy sequences. Ideally, these peptides should be experimentally validated and matched in terms of length and physicochemical property distributions relative to the active set. When experimentally confirmed inactive peptides are not available, non-AMP sequences or synthetic decoys may be used; however, this approach introduces uncertainty, as such sequences may reflect a lack of annotation rather than true inactivity. Thus, a balanced and well-curated dataset is essential to enable robust model training and to support the identification of meaningful sequence–activity relationships underlying antifungal properties [51].

Building on this curated dataset, the next stage corresponds to descriptor calculation (2), which constitutes the core of the QSAR approach. This step establishes the quantitative link between peptide

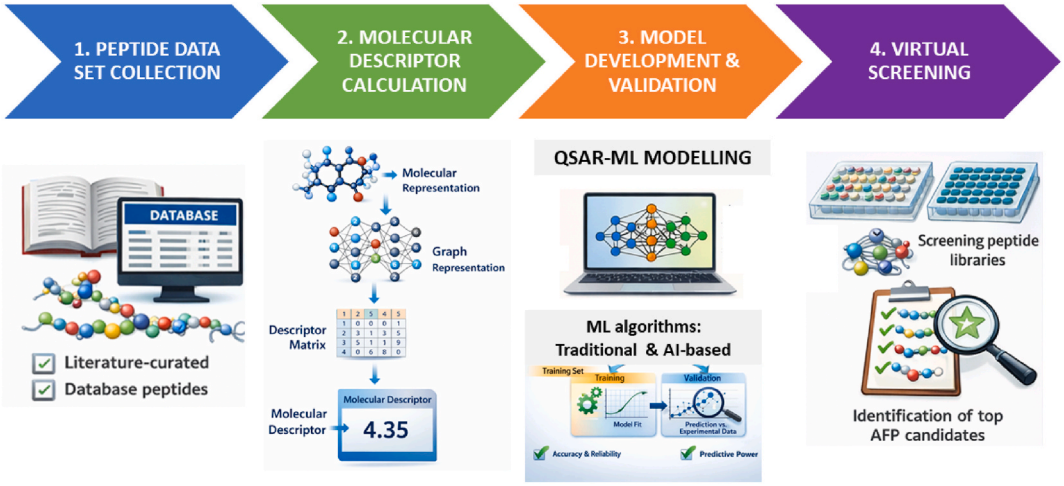


**Fig. 2.** Conceptual QSAR and ML workflow for antifungal peptide discovery. ChatGPT Pro was used to support the graphical organization and layout of the workflow diagram.

structure and antifungal activity by extracting chemical and structural information. A variety of software tools can be employed to calculate molecular descriptors, which fall into three main categories: (i) classical 2D and 3D molecular descriptors (e.g., physicochemical, topological); (ii) sequence-based descriptors (e.g., amino acid composition); and (iii) novel machine learning-derived embedding descriptors, which represent high-dimensional vectors capturing complex sequence, structural, and functional patterns. This process of feature engineering transforms raw peptide sequences into structured numerical representations suitable for downstream modeling.

Once the feature set is defined, the workflow proceeds to the model development phase (3). This stage involves feature selection to identify the most informative descriptors, followed by algorithm selection from a range of machine learning techniques. The modeling toolkit includes classical algorithms such as support vector machines (SVM), random forest (RF), and linear regression, as well as more advanced deep learning approaches, including artificial neural networks (ANN), recurrent neural networks (RNN), and generative adversarial networks (GANs), such as feedback GANs (FBGAN). Subsequently, model training and validation are performed to ensure robustness and predictive performance. The resulting models are rigorously evaluated using internal and external validation strategies, including k-fold cross-validation, leave-one-out cross-validation (LOOCV), or hold-out train/test splits. Model performance is then quantified using metrics such as $R^2$, accuracy, and AUC to ensure reliability prior to deployment [52,53].

Finally, the last phase involves the virtual screening of peptide dataset to identify potential antifungal peptides candidates (4). This selection is guided by information encoded in molecular descriptors and ML-QSAR models, which provide interpretable insights through feature importance analysis, highlighting the key peptide properties that drive activity. The resulting prioritized candidates support the rational design of stable, potent, and cost-effective peptides tailored for antifungal applications. By enabling sequence- and motif-level screening in a predictive framework, this approach accelerates discovery and facilitates structure-informed optimization of next-generation AFPs [34].

### 2.1. Evolution of QSAR and ML in antifungal peptide discovery

As described in Section 1.4, a structured literature search was conducted to identify and analyze studies published over the last decade (2015-2025) that have applied QSAR and machine learning approaches to antifungal peptide discovery. The selected studies are summarized in Tables 2A and 2BA and 2B, each study is systematically characterized according to key criteria, including bibliographic reference; dataset composition used for model training and validation; modeling type (classification or regression); predicted biological endpoints, such as antifungal activity, minimum inhibitory concentration (MIC), half-maximal inhibitory concentration ($IC_{50}$), pMIC, or $pIC_{50}$; descriptor type (e.g., classical QSAR descriptors, sequence-based features, or machine learning-derived embeddings); modeling algorithms (e.g., random forest, support vector machine); performance metrics (e.g., $R^2$, $Q^2$); validation strategy (e.g., cross-validation, external test sets); target fungal species; and software used for descriptor calculation (e.g., RDKit, Dragon).

By analyzing the studies reported in Tables 2A and 2BA and 2B, it is evident that the application of QSAR modeling and machine learning to antifungal peptide design has diversified substantially over the past decade. This diversification is reflected in the coexistence of approaches ranging from small, chemically constrained datasets combined with interpretable linear QSAR models, and large-scale, multi-species datasets integrated with advanced machine learning and generative AI pipelines. Accordingly, classical methodologies remain widely used and are progressively complemented, rather than replaced at present, by ML- and AI-based strategies.

From a methodological perspective, the field includes small QSAR studies based on classical molecular descriptors, larger classification models relying on sequence- or physicochemical-based features, approaches incorporating protein-language-model embeddings, and fully generative design pipelines. While small regression datasets typically support interpretable, local structure–activity relationship analyses, larger classification frameworks are more suitable for broad screening and candidate prioritization. Thus, the coexistence of large-scale predictors and smaller interpretable models reflects the fragmentation and specialization of the field, not a simple replacement of classical QSAR by deep learning.

Considering that the robustness of any QSAR and ML model is fundamentally determined by the quality and characteristics of its training data, this aspect warrants careful analysis. Early QSAR studies based on small datasets ($N < 100$), such as those by Borkar et al., and Yousefinejad et al., were limited in scale but benefited from relatively homogeneous peptide series and well-defined descriptor spaces [54,55]. Under these conditions, linear or semi-linear models can perform effectively, as they primarily interpolate within a restricted chemical domain rather than extrapolating across broad peptide diversity. This contextualizes the strong external performance reported for the SPCR model by Yousefinejad et al., which achieved a high $R^2$ test despite being trained on only 60 peptides [55]. Importantly, such results should not be interpreted as evidence that small datasets are inherently superior, but rather as a consequence of carefully curated descriptor selection and a constrained applicability domain that enables robust local modelling. These approaches remain highly valuable for mechanistic interpretation and lead optimization, although their generalizability beyond the original peptide space is inherently limited.

A similar rationale applies to the study by Chang et al., where a

**Table 2A**
QSAR and machine learning studies on antifungal peptides (2015-2025): study design, descriptors, algorithms, target species, and dataset structure.

| N total (Train/Test/Ext) | Endpoint | Descriptors | Model | Species | Ref |
|---|---|---|---|---|---|
| 45 (34/–/11) | pMIC | Seq + 2D | G/PLS | *C. albicans* | [54][a] |
| 60 (48/–/12) | $pIC_{50}$ | 3D | SPCR | *C. albicans* | [55] |
| 17514 (1556/430/15528) | Class. (AFP/Non-AFP) | Seq | RF | General antifungal | [56] |
| 11550 (Class.); 1583–95–275–148 (Reg.) | Class. + pMIC | 2D | SVC/SVR | *Candida* spp. | [42] |
| 149 (18/–/149) | MIC (μg/mL) BSI | Seq + 2D + Exp | PLSR | *Candida* spp. | [57] |
| 2488 (1990/–/498) | Class. (AFP/Non-AFP) | Seq + 2D | SVC, kNN, RF, LR, MLP, Stack | Multiple fungi[b] | [58] |
| 294 | Class. | Seq + 2D | SVC | General antifungal | [59] |
| 8258 (5506/1376/1376) | Class. (AFP/Non-AFP) | Emb (ESM-2) | BERT-MLP | *Candida* spp. | [60] |

Abbreviations: MIC, minimum inhibitory concentration; pMIC, negative logarithm of MIC; $IC_{50}$, half-maximal inhibitory concentration; $pIC_{50}$, negative logarithm of $IC_{50}$; AFP, antifungal peptide; Seq, sequence-based descriptors; 2D, traditional 2D QSAR descriptors; 3D, 3D QSAR descriptors; Emb, embedding descriptors; RF, Random Forest; SVC, Support Vector Classifier; SVR, Support Vector Regression; BSI, Broad-Spectrum Selectivity Index; PLSR, Partial Least Squares Regression; SPCR, Principal Component Regression; MLP, Multilayer Perceptron.

[a] Rows correspond to the same studies and reference order shown in Table 2B.

[b] Fungal species included: Candida, Aspergillus, Fusarium, Botrytis, Cryptococcus, Saccharomyces, Pichia, Batrachochytrium, Neurospora, Didymella, Leptosphaeria, Phytophthora, Verticillium, Fulvia and Alternaria.

**Table 2B**
QSAR and machine learning studies on antifungal peptides (2015–2025): predictive performance and validation strategy.

| Metric type | Performance | Validation | Ref |
|---|---|---|---|
| Regression | $R^2 = 0.85$; $Q^2 = 0.75$; $R^2$pred = 0.50 | LOO, LSO, bootstrap, Y-scrambling; external test | [54][a] |
| Regression | $R^2 = 0.97$; $Q^2$cv = 0.82; $R^2$test = 0.87 | CV, Y-scrambling; external test | [55] |
| Classification | Train: ACC = 0.95, MCC = 0.91, AUC = 0.99; Test: ACC = 0.94, MCC = 0.88, AUC = 0.99; External: ACC = 0.86, MCC = 0.66, AUC = 0.91 | 10-fold LSO CV; external test | [56] |
| Classification | ACC = 0.89; MCC = 0.79; AUC = 0.95 | 10-fold CV | [42] |
| Regression | $R^2$train = 0.93-0.94; $R^2$val = 0.66–0.89 | 10-fold CV | [42] |
| Regression | $R^2 = 0.734$; prediction accuracy within one dilution = 78.2% | LOO; repeated 9-fold CV | [57] |
| Classification | Train: ACC = 0.997, MCC = 0.995; Test: ACC ≈ 0.88, MCC ≈ 0.75, F1 ≈ 0.88 | 5-fold CV; independent test | [58] |
| Classification | Model 1: ACC = 0.88, MCC = 0.76; Model 2: ACC = 0.90, MCC = 0.80 | Not specified | [59] |
| Classification | Validation: ACC = 0.89, MCC = 0.79; Test: ACC = 0.96, MCC = 0.91 | Internal validation; independent test | [60] |

ACC, accuracy; MCC, Matthews correlation coefficient; AUC, area under the receiver operating characteristic curve; F1, F1-score; CV, cross-validation; LOO, leave-one-out cross-validation; LSO, leave-some-out or leave-subject-out validation, depending on the original study; $R^2$, coefficient of determination; $Q^2$, cross-validated coefficient of determination; $R^2$pred, predictive coefficient of determination on an external set.

[a] *Rows correspond to the same studies and reference order shown in* Table 2A.

relatively small dataset and a PLSR-based framework were employed not as a regression to simpler methodologies, but as a deliberate modelling choice aimed at deriving interpretable and quantitative structure-activity relationships linking peptide features, antimicrobial activity, and mammalian cytotoxicity [57].

By contrast, larger classification datasets, such as those used by Zhang et al., Pinacho-*Castellanos* et al., Lobo et al., and Yin et al., enable broader virtual screening and improved generalization across diverse peptide sequences, with dataset sizes ranging from several thousand peptides to more than 17,000 sequences depending on the study design [42,56,58,60]. These models typically report strong predictive performance, with validation or test accuracies ranging from approximately 0.86 to 0.96 and AUC values up to 0.99. However, such performance does not scale linearly with dataset size. Once a sufficient data volume is reached to train stable classifiers, further improvements appear to depend more critically on data quality than on quantity, including rigorous curation, redundancy reduction, reliable construction of negative sets, endpoint consistency, and strict separation between training and external validation data. This aspect is particularly relevant in antimicrobial peptide modelling, where inactive sequences are frequently inferred from the absence of annotation rather than experimentally confirmed inactivity.

The comparison between classical QSAR descriptors and protein-language-model embeddings further highlights a central methodological trade-off between interpretability and representational power. Classical 2D, 3D, physicochemical, and sequence-based descriptors preserve a higher degree of chemical interpretability, enabling the identification of properties such as hydrophobicity, charge distribution, amphipathicity, molecular size, and spatial organization, which are directly relevant for rational peptide optimization [54,55,57]. This is evident in descriptor-based QSAR studies such as those by Borkar et al. and Yousefinejad et al., where relatively small and chemically constrained peptide datasets enabled the extraction of interpretable local structure–activity relationships [54,55]. Similarly, the PLS-based framework reported by Chang et al. illustrates how experimentally accessible physicochemical properties, including RP-HPLC retention time, can be incorporated into transparent models aimed at optimizing potency and selectivity within focused peptide series [57]. Such interpretability is especially valuable for agricultural translation, where model outputs must inform not only antifungal potency but also stability, phytotoxicity, solubility, persistence, and formulation compatibility. However, the requirements for broad virtual screening and *de novo* peptide discovery differ from those of local QSAR optimization. Larger and more heterogeneous datasets, such as those used by Zhang et al. and Pinacho-Castellanos et al., support wider candidate prioritization but also shift the modelling objective from mechanistic explanation toward classification performance and generalization across diverse peptide spaces [42,56]. In this context, non-linear machine learning algorithms and high-dimensional feature representations become increasingly useful, although the resulting structure-activity rules are less direct. Lobo et al. further illustrate that predictive performance depends on the interaction between dataset structure, feature representation, dimensionality reduction, and algorithm selection, rather than on a universally superior modelling approach [58]. Protein-language-model embeddings and deep learning architectures, exemplified by the ESM-derived representations and neural-network classifiers used by Yin et al., extend this trend by capturing complex and non-local sequence relationships that are difficult to encode through predefined descriptors [60]. These methods are therefore highly promising for large-scale screening and generative peptide design, but their reduced interpretability remains a major limitation when predictions must be converted into explicit design rules for crop-protection applications.

Endpoint definition represents another critical determinant of model utility. Regression approaches based on MIC, pMIC, $IC_{50}$, or selectivity indices provide quantitative resolution for potency optimization but require highly consistent experimental protocols and are sensitive to inter-laboratory variability [42,54,55,57]. In contrast, binary classification frameworks are more scalable and effective for large-library prioritization; however, they reduce biological activity to simplified active-inactive labels, thereby losing information on potency gradients, spectrum of activity, and dosage requirements [56,58–60]. This limitation is particularly relevant for agricultural antifungal peptide (AFP) discovery, where field performance cannot be inferred solely from *in vitro* classification outcomes.

Validation strategy is also central to interpreting these reported performances. Internal validation procedures, such as leave-one-out, k-fold cross-validation, or bootstrap resampling, are useful for estimating model stability within a defined dataset, but they do not necessarily demonstrate generalizability beyond the original peptide space. In this respect, external test sets, independent validation, similarity-reduced splits, and applicability-domain assessment are more informative for evaluating whether a model can support prospective candidate prioritization [50,54–60]. This distinction is particularly important for agricultural AFP discovery, because models trained and validated mainly on biomedical or Candida-centered datasets cannot be assumed to transfer directly to phytopathogenic fungi or field-relevant endpoints without plant-pathogen-specific external validation.

Taken together, the reviewed studies indicate that dataset size, descriptor representation, and model complexity must be aligned with the intended predictive objective. Small, interpretable QSAR models remain valuable for local structure-activity exploration and lead optimization, whereas larger ML classifiers are more appropriate for broad screening and prioritization tasks. Protein-language-model and generative approaches further expand the chemical space accessible to computational design but require rigorous external validation and complementary interpretability frameworks. Accordingly, the most effective strategy for agricultural AFP discovery is not the replacement

of classical QSAR methodologies by deep learning, but their integration within hybrid frameworks that combine interpretable descriptors, robust validation strategies, applicability-domain assessment, and scalable ML architectures.

Importantly, the analysis of the literature reveals that the vast majority of existing studies (Tables 2A and 2BA and 2B) focus on antifungal peptide discovery in biomedical or clinically oriented contexts. This predominance highlights a translational gap rather than a methodological limitation, as QSAR and machine learning approaches are well established but remain comparatively underexplored in agricultural and agri-food applications. Only a limited number of studies explicitly address antifungal peptide prediction in plant-pathogen or crop-protection contexts, underscoring a significant opportunity for future research. Extending these computational frameworks to agricultural systems therefore represents a promising avenue to bridge this gap and enable rational peptide design for plant disease control.

## 3. Data resources for QSAR and ML antifungal peptide model development

Although QSAR and machine learning approaches for antifungal peptide (AFP) discovery have been primarily driven by advances in algorithms and molecular representations, their development critically depends on parallel progress in data availability, quality, and curation. In fact, dataset size, diversity, and reliability are key determinants of model performance, generalizability, and applicability domain. Over time, the field has evolved from small, manually curated datasets to large-scale, publicly available peptide repositories. Consequently, a clear understanding of the origin, composition, and intended use of the databases employed for model training and validation is essential for interpreting results, ensuring reproducibility, and enabling fair benchmarking across studies.

Across the reviewed literature, several databases stand out as central resources for constructing training datasets. DRAMP (Data Repository of Antimicrobial Peptides), APD3 (Antimicrobial Peptide Database 3), CAMP (Collection of Anti-Microbial Peptides), and DBAASP (Database of Antimicrobial Activity and Structure of Peptides) are the most frequently used sources of antimicrobial and antifungal peptides. These databases are often combined to generate larger and more chemically diverse datasets, as summarized in Table 3.

This multi-source strategy, applied in studies such as those by Zhang et al., Lobo et al., Medvedeva et al., and Yin et al., reflects the need to maximize coverage of known active peptides while mitigating biases inherent to individual repository curation policies [42,58–60]. Importantly, DRAMP and DBAASP not only provide peptide sequences but also include experimentally derived quantitative activity data (e.g., MIC values) and, in some cases, structural information, which makes them particularly valuable for QSAR regression modelling.

In addition, StarPepDB has emerged as a widely used meta-database that integrates experimentally annotated peptide information from multiple primary sources, including UniProtKB, APD, and YADAMP [56,60]. Studies such as those by Pinacho et al., and Yin et al., have used StarPepDB as a central resource for dataset construction and model evaluation. The specific role of each database across QSAR and ML workflows is further detailed in Table 4, which highlights how different data sources are combined depending on the modelling objective (classification vs regression, antifungal vs general antimicrobial prediction).

In contrast, earlier QSAR studies, such as those by Borkar et al., and Yousefinejad et al., relied mainly on manually curated datasets extracted from the scientific literature rather than large integrated repositories [54,55]. Although more limited in size, this strategy allowed precise control over data quality and composition, as well as direct linkage to well-characterized experimental evidence. As a result, they were particularly effective for deriving robust structure–activity relationships within narrow chemical spaces, albeit with a more limited applicability domain.

A particularly important aspect of dataset construction is the definition of negative or decoy sequences for classification tasks. In most QSAR and ML studies (e.g., Yin et al., and Pinacho et al.), UniProt has been used as the primary source of non-antimicrobial peptide sequences (Tables 3 and 4) [56,60]. The extensive coverage and diversity of UniProt help reduce sampling bias in negative datasets, thereby improving classifier robustness. In addition, UniProt underpins modern protein language models, further reinforcing its importance in contemporary peptide prediction pipelines.

However, integrating multiple databases also introduces non-trivial curation challenges. Meta-databases such as StarPepDB increase

**Table 3**
Overview of antimicrobial peptide databases and their application in machine learning and QSAR studies.

| Database | Focus | Key Features & Usage Context | Relevance & Application in the Studies | References |
|---|---|---|---|---|
| **DRAMP** (Data Repository of Antimicrobial Peptides) | Therapeutic and experimental AMPs | Used in multiple studies for building ML predictors. Noted for extensive annotations and being a primary source for curated, high-quality data. | **High.** A widely used curated source for assembling AMP datasets, although additional deduplication, annotation harmonization, and activity-label verification are required before model training. | [42, 58–60] |
| **APD3** (Antimicrobial Peptide Database 3) | Natural and synthetic AMPs | Frequently used in combination with DRAMP, CAMP, and DBAASP to build large, diverse datasets for model training and virtual screening. | **High.** Another core database used in multi-source consolidations to ensure dataset comprehensiveness and reduce individual database bias. | [42,59,60] |
| **DBAASP** (Database of Antimicrobial Activity and Structure of Peptides) | AMP activity and structure | Explicitly mentioned for containing quantitative activity data (e.g., MICs). Used in studies focusing on structure-activity relationships. | **High.** Particularly valuable for QSAR-related work due to its structured activity data, which is crucial for building quantitative models. | [42, 58–60] |
| **CAMP** (Collection of Anti-Microbial Peptides) | Antimicrobial Peptides | Consistently used as part of the "quartet" of core databases (with DRAMP, APD3, DBAASP) for model development and benchmarking. | **High.** Its well-established and curated nature makes it a reliable source for assembling large-scale AMP datasets. | [42,59,60] |
| **UniProt** (incl. Swiss-Prot) | General protein knowledgebase | The nearly universal source for non-AMP sequences (negative controls) in machine learning studies. Also the underlying data source for pre-trained protein language models (ESM1b, ProtTrans). | **Critical (for a specific purpose).** Indispensable for sourcing negative data and for next-generation model development via protein Language Models (pLMs). | [42,56,60] |
| **StarPepDB** | AMPs (a meta-resource) | Described as a database compiled from other primary sources (UniProtKB, YADAMP, APD, etc.). Used as a comprehensive starting point for data collection. | **Medium/High.** Efficient for initial data mining as it aggregates several sources, but its derivative nature means it may inherit biases. | [56,60] |
| **Literature/Specialized Compilations** | Varies (e.g., specific peptide families) | Used for manual curation of specific, well-characterized peptide sets. Allows for precise control and connection to established experimental work. | **Variable.** High for specific, legacy data but not scalable for large ML projects. | [54,55] |

**Table 4**
Role of databases and data sources used in QSAR and ML-based antifungal peptide studies reported [42,54–60].

| Database(s) Used | Role in the Study | Reference |
|---|---|---|
| **Literature** | **Training & Testing:** This study did not use a public database. The dataset of protegrin peptides was a compilation of sequences and activities from numerous experimental studies. | [54] |
| | **Training & Testing:** This study did not use a public database. The dataset of 60 synthetic hexapeptides and their experimental activities against *C. albicans*, MRSA, MRSE, and *E. coli* was taken directly from a single research paper (Sharma et al., 2010). | [55] |
| **StarPepDB** | **Training:** The primary source for building all training, test, and external validation sets. The study explicitly uses the "largest experimentally validated nonredundant peptide data set reported to date," which is a compiled database called StarPepDB. | [56] |
| **Various, via benchmarking** | **Evaluation/Benchmarking:** For comparison, the study uses peptides from the training sets of 13 other predictors (like ClassAMP, iAMP-2L, AntiFP, CAMPR3, etc.) to create unbiased benchmarking datasets. | |
| **DBAASP, APD3, DRAMP, CAMP** | **Training:** Used to compile Data set 1 (for classification) and Data set 2 (for activity prediction). | [42] |
| **UniProt Knowledgebase** | **Screening:** A library of over 3 million peptides was collected from UniProt to serve as the initial pool for virtual screening. | |
| **DBAASP** | **Training:** The entire dataset for training and testing the AFP prediction models was sourced exclusively from DBAASP, based on quantitative MIC values. | [58] |
| **DBAASP** | **Training:** Used to curate the dataset of peptides with confirmed single-action or triple-action activity, relying on its quantitative MIC/IC data. | [59] |
| **DRAMP & APD3** | **Training/Design:** Used to analyze 1237 anti-Candida peptides to establish "dominant amino acids" and physicochemical property thresholds for generating candidate peptides. | [60] |
| **BLAST** | **Screening:** Used to filter out generated peptides with high similarity to known sequences, ensuring novelty. | |

dataset coverage but may also propagate redundancy, duplicated entries, inconsistent annotations, and biases inherited from their source databases. Therefore, rigorous preprocessing steps - such as sequence deduplication, homology reduction, harmonization of activity labels, and explicit tracking of data provenance - are essential prior to model training. Similarly, while UniProt is widely used to generate negative datasets, this approach is not without limitations, as some sequences labeled as negatives may simply lack experimental antifungal evidence rather than being truly inactive, introducing potential false negatives. This issue is particularly relevant in antifungal peptide prediction, where experimentally validated inactive peptides remain scarce. Best practices therefore include the use of non-redundant datasets, similarity thresholding between training and test sets, exclusion of poorly annotated sequences, construction of decoy sets matched in length and physicochemical properties, and, whenever possible, external validation using experimentally confirmed active and inactive peptides against defined fungal targets.

Overall, the evidence across the reviewed studies suggests that no single database is inherently sufficient or limiting. Instead, predictive performance and generalizability depend on the strategic integration of multiple data sources, appropriate curation pipelines, and the alignment between dataset design and modelling objectives. Notably, variability in predictive performance across studies using similar databases indicates that limitations often arise not from data origin, but from model design choices, descriptor quality, and validation strategies Tables 2A, 2B, 3, and 4

A comparative analysis of the reviewed studies (Tables 2A, 2B, 3, and 4) reveals consistent patterns in data sourcing and model performance. Quantitative QSAR models of antifungal activity typically rely on literature-derived datasets and public repositories such as DRAMP, APD3, CAMP, UniProt, and DBAASP. Interestingly, the model developed by Yousefinejad et al., achieved strong generalization performance despite being trained on a relatively small manually curated dataset [55], highlighting the continued relevance of carefully designed traditional QSAR approaches. In contrast, classification models generally follow two main strategies: one integrates UniProt and StarPepDB (as in Pinacho-Castellanos et al. and Yin et al., the latter also incorporating additional sources), achieving high discriminative performance on external datasets [56,60]; the other combines DRAMP and DBAASP with complementary databases, as reported by Lobo et al., Medvedeva et al., and Yin et al., also yielding robust classification results [58–60].

Nevertheless, the suitability of a database cannot be evaluated solely based on downstream predictive performance. In practice, both data quality and model design jointly determine outcomes, and different combinations of databases and modelling strategies can yield models ranging from moderate to high discriminative power.

Finally, a key limitation of currently available peptide resources is their incomplete coverage of plant-pathogenic fungi, the lack of standardized quantitative activity measurements under agriculturally relevant conditions, and heterogeneous annotation standards. Although DRAMP, APD3, DBAASP, CAMP, StarPepDB, and related resources provide a solid foundation for AFP modelling (see Tables 3 and 4), these limitations highlight the need for dedicated, standardized datasets specifically designed for plant-pathogen antifungal peptide research to support reliable QSAR and ML applications in crop protection.

## 4. Future perspectives and challenges in QSAR and ML-driven antifungal peptide discovery for crop protection

The reviewed literature highlights a pronounced translational gap between antifungal peptide prediction pipelines developed in biomedical research and their potential application in crop protection and the agro-food sector. To date, most QSAR and ML models have been trained and validated predominantly using datasets derived from clinically relevant fungal pathogens, particularly *Candida* spp. This biomedical orientation has driven substantial methodological progress; however, the reviewed literature shows limited systematic exploration of plant-pathogen systems and, consequently, limited development of AFPs tailored to agricultural applications.

A critical observation is that most QSAR and ML developments in antifungal peptide discovery over the past decade have focused on antimicrobial peptides targeting human pathogens [61–63]. This predominance does not reflect an intrinsic limitation of the computational methodologies, but rather the absence of systematic cross-domain validation under plant-pathogen-specific conditions. Accordingly, the current gap should be interpreted not as evidence of poor transferability, but as an underexplored opportunity to extend established computational frameworks toward agro-food applications.

From a mechanistic standpoint, such transferability is supported by significant biological conservation between human- and plant-pathogenic fungi. Both systems share key structural components, including chitin-rich cell walls, β-glucans, mannoproteins, and membrane sterols. Consistently, antifungal peptides often act through conserved mechanisms such as membrane permeabilization, cell-wall disruption, intracellular targeting, induction of oxidative stress, and modulation of host defense responses [20,22–28]. These shared features provide a solid conceptual foundation for leveraging QSAR and ML models originally developed in clinical contexts for agricultural applications.

However, the transition toward crop protection introduces important context-dependent constraints. In contrast to controlled clinical environments, plant-associated ecosystems are highly heterogeneous and dynamic, encompassing the phyllosphere, rhizosphere, fruit surfaces, and post-harvest storage systems. In these environments, peptide performance is strongly modulated by abiotic and biotic factors, including ultraviolet radiation, temperature and humidity fluctuations, pH variability, proteolytic degradation, adsorption to plant cuticular waxes, and formulation-dependent delivery effects [20,39].

Consequently, antifungal peptide design for agricultural applications must extend beyond intrinsic antifungal potency and incorporate agro-relevant performance attributes, including environmental stability, persistence, delivery efficiency, phytotoxicity, and selectivity toward beneficial plant-associated microbiota [20,39]. In this regard, existing QSAR and ML pipelines and peptide-design strategies derived from biomedical research constitute a robust foundation that can be systematically adapted to the specific constraints of plant-pathogen systems [36,42].

Importantly, this observation should not be interpreted as a limitation of the methodology itself, but rather as an underexploited opportunity to extend established computational frameworks toward agro-food-relevant antifungal discovery. Indeed, the transferability of QSAR and ML approaches is supported by mechanistic commonalities between fungal systems; however, their application in agriculture requires careful consideration of additional biological, environmental, and formulation-related constraints that are absent or less prominent in clinical settings [20,39].

For crop-protection-oriented applications, model development should explicitly integrate plant-pathogen relevance and agro-environmental descriptors. This includes not only bioactivity against phytopathogens but also environmental and formulation-related properties that govern field performance. Unlike clinical settings, plant-associated environments are subject to complex and fluctuating conditions that significantly influence peptide stability and efficacy. Therefore, predictive frameworks must move beyond intrinsic activity and incorporate variables governing real-world applicability, including stability, delivery, persistence, and safety under plant-associated conditions [36,39,42].

From a validation perspective, current QSAR and ML workflows require more stringent and domain-specific evaluation protocols. External validation sets should be enriched with agriculturally relevant phytopathogens, such as *Botrytis cinerea*, *Fusarium graminearum*, *Magnaporthe oryzae, Alternaria* spp., and *Podosphaera xanthii*. Experimental studies on antifungal peptides active against plant-pathogenic fungi, such as the peptide P852 against *Fusarium oxysporum*, illustrate the relevance of validating AFP candidates directly in plant-pathogen systems [64]. Moreover, computational predictions should be complemented by hierarchical experimental validation strategies that progressively approximate field conditions, ranging from *in vitro* assays to plant tissue systems, detached-leaf assays, greenhouse trials, and formulation-based evaluations. Such multi-level validation frameworks are essential to assess whether predicted AFPs retain their activity, stability, and functional performance under realistic agricultural conditions [39,64].

Overall, while existing QSAR and ML frameworks and biomedical peptide-design paradigms provide a robust and well-established methodological foundation, their translation to agriculture remains limited. This review therefore identifies a clear and timely opportunity to adapt these computational strategies toward the systematic discovery of antifungal peptides for crop protection and agro-food applications, supported by domain-specific datasets, agro-oriented descriptors, and standardized validation pipelines tailored to plant-pathogen interactions.

### *4.1. Curating plant-pathogen specific datasets: The cornerstone of predictive power*

The performance of any QSAR and ML model is fundamentally constrained by the quality, consistency, and representativeness of the data on which it is trained [50,51]. As highlighted in Table 4, current models rely heavily on databases such as DRAMP, APD3, CAMP, StarPepDB, UniProt, and DBAASP, which have been extensively used for antimicrobial and antifungal peptide prediction but were not specifically developed to capture plant-pathogen-specific endpoints or field-relevant crop-protection conditions [42,56,58,60]. Accordingly, although these resources provide a valuable foundation for AFP modelling, they offer limited standardized representation of agriculturally relevant phytopathogenic fungi, including *Botrytis cinerea, Fusarium graminearum, Magnaporthe oryzae, Alternaria* spp., and *Podosphaera xanthii* [7–9,20,39]. The first and most critical step toward bridging this gap is therefore the systematic curation of high-quality, plant-pathogen-specific peptide interaction datasets.

This effort must go beyond simple sequence aggregation. Future datasets should be meticulously annotated with quantitative activity data, such as MIC, $IC_{50}$, percentage inhibition of mycelial growth, or inhibition of spore germination, against defined panels of phytopathogenic fungi and measured under standardized conditions to minimize inter-laboratory variability [42,50,51]. Crucially, these datasets must be carefully balanced, including not only active peptides but also well-defined inactive or weakly active decoy sequences. This is essential for training robust classifiers capable of distinguishing true antifungal activity from non-functional peptides and for avoiding biases derived from negative sets based only on missing annotations [50,51,56,60]. Furthermore, these resources should be made publicly available to foster community-wide benchmarking and accelerate the development of next-generation QSAR and ML models specifically tailored for agrochemical and crop-protection applications, an approach that has proven foundational in biomedical peptide research [50,56].

As a concrete example, the community could develop a curated benchmark dataset focused on peptide activity against a defined panel of major phytopathogenic fungi, such as *Botrytis cinerea*, *Fusarium graminearum*, *Magnaporthe oryzae*, *Alternaria* spp., and *Podosphaera xanthii* [7–9]. Such a dataset should include standardized quantitative endpoints, including MIC, $IC_{50}$, or percentage inhibition of mycelial growth or spore germination, measured under harmonized experimental conditions. Ideally, it should also incorporate inactive or weakly active peptides, as well as relevant molecular and biophysical descriptors, including peptide length, net charge, hydrophobicity, predicted secondary structure, and protease susceptibility [29,30,50,51]. Additionally, when available, agro-relevant properties such as phytotoxicity, formulation compatibility, leaf-surface retention, and environmental stability should be included [20,39]. A benchmark of this type would enable direct comparison of descriptor sets, machine learning algorithms, applicability domains, and external predictive performance in a plant-pathogen-specific context [50,56].

### *4.2. Developing standardized QSAR and ML frameworks for agricultural relevance*

The creation of specialized datasets must be accompanied by the development of robust and rigorously validated QSAR and ML frameworks for agriculturally relevant targets. The architectural diversity seen in Tables 2A and 2BA and 2B, from interpretable PLS models [57] to complex pLM-based classifiers [60], provides a rich toolkit, but its application to plant pathogens is currently nascent.

Future research should systematically assess the relative performance and transferability of different algorithmic families, descriptor spaces, and validation strategies for predicting antifungal activity against phytopathogens exhibiting diverse cell wall architectures and infection strategies. In particular, it remains unclear whether low-

complexity, interpretable models based on physicochemical descriptors (e.g., hydrophobicity, charge, and amphipathicity) are sufficient to capture broad-spectrum antifungal activity, or whether deep learning approaches are required to resolve higher-order sequence–structure–activity relationships, especially for host-adapted and species-specific fungal interactions.

Within this framework, the establishment of community-accepted benchmark datasets, standardized evaluation protocols, and harmonized performance metrics for plant-pathogen models is essential. Such frameworks would enable direct and reproducible comparison across modelling strategies, facilitate objective assessment of model generalizability, and ultimately support a more rational and evidence-driven selection of QSAR and ML approaches for antifungal peptide discovery in agricultural systems.

### 4.3. Integrating generative modeling with QSAR screening for rapid candidate prioritization

Future antifungal peptide (AFP) discovery should extend beyond the prediction of known active sequences toward the design of candidates optimized for defined biological and physicochemical requirements. Recent advances in generative modeling, including AMPd-Up, ProT-Diff, Fung-AI, and FBGAN-based architectures, illustrate the capacity of deep learning frameworks to explore the peptide sequence space and generate novel antimicrobial and antifungal candidates [46–49].

A transformative direction for the coming decade is the tight integration of generative models with QSAR-based screening into a unified, multi-stage discovery pipeline. In this paradigm, generative algorithms trained on plant-pathogen-relevant datasets would be used to produce large virtual libraries of *de novo* peptide sequences. These libraries would then be subjected to rapid, high-throughput QSAR filtering to prioritize candidates with high predicted antifungal activity, following the broader logic of QSAR-guided peptide prioritization and virtual screening [34,42,50].

Subsequently, top-ranked sequences could be refined through secondary *in silico* screening layers aimed at predicting key developability and agro-relevant properties, including low toxicity, improved solubility, and enhanced protease stability [29,30,39]. This hierarchical prioritization framework would enable the identification of compact, high-confidence candidate sets for experimental validation.

Under this framework, the iterative coupling of computational generation, QSAR-based filtering, and experimental feedback has the potential to accelerate the discovery cycle and enable the rational design of crop-specific biofungicides tailored to defined plant diseases [39, 46–50].

### 4.4. Incorporating field-relevant traits into multi-objective design

As outlined above, a fundamental distinction between biomedical and agricultural peptide design lies in the nature of the target environment. In therapeutic applications, peptides are typically optimized for potency, biological stability, delivery, and reduced cytotoxicity toward mammalian cells [30,34,57]. By contrast, in crop protection, antifungal peptides must retain functionality on plant surfaces, where they are exposed to highly variable and stress-intensive conditions, including ultraviolet radiation, rainfall, temperature and humidity fluctuations, endogenous and microbial proteases, and interactions with the hydrophobic plant cuticle [20,29,30,39].

This requirement should not be interpreted as a simple maximization of peptide stability, but rather as the optimization of a controlled persistence–degradation trade-off, whereby AFPs remain active for a sufficient time to suppress pathogen development while subsequently undergoing environmentally benign degradation to minimize accumulation and off-target effects [20,30,39].

Accordingly, next-generation QSAR and ML frameworks should transition from single-objective activity prediction toward multi-objective optimization paradigms explicitly tailored to agricultural deployment. Beyond antifungal potency against relevant phytopathogens, predictive models should incorporate key field-relevant descriptors and endpoints, including protease resistance, UV stability, aqueous solubility, leaf-surface retention, cuticle permeability, formulation compatibility, and phytotoxicity [29,30,39,50]. These parameters are essential to ensure that candidate peptides are not only active under *in vitro* conditions but also functionally viable under realistic crop environments.

To achieve this, future modelling strategies should integrate interpretable physicochemical descriptors with learned sequence representations, including embedding-based features, enabling simultaneous optimization across biological activity, stability, delivery efficiency, and safety constraints [34,50,58,60]. However, the absence of harmonized descriptor frameworks and standardized agricultural benchmarking datasets currently limits cross-study comparability and model generalizability, highlighting the need for curated resources specifically designed for crop-protection-oriented peptide engineering [50,56].

At present, most AFP prediction systems remain largely restricted to binary classification, such as active/inactive prediction, or potency estimation under controlled laboratory conditions, such as MIC or pMIC prediction [42,56,58,60]. Critical field-relevant properties, including environmental degradation kinetics, residual activity after UV exposure, rainfastness, surface adhesion, cuticular transport, and compatibility with encapsulation or adjuvant systems, are rarely incorporated into model training or validation. Consequently, the development of future agricultural AFP datasets should systematically include these experimentally measurable endpoints, enabling QSAR and ML frameworks to evolve from purely bioactivity predictors into decision-support tools for the rational design of deployable crop-protection agents [39,50,56].

## 5. Conclusions

QSAR and machine learning approaches have significantly advanced the field of antifungal peptide discovery; however, their direct translation into crop protection applications remains limited. The body of evidence reviewed in this work indicates that current predictive frameworks are largely rooted in biomedical data, predominantly focused on *Candida* spp., and typically validated using endpoints that capture intrinsic antifungal potency under simplified *in vitro* conditions. As a result, the extension of these models to agricultural contexts cannot be considered straightforward, but rather requires explicit adaptation and validation within plant-pathogen-specific systems.

In this sense, the main limitation is not the methodological maturity of QSAR and ML itself, but the mismatch between the data and validation paradigms used in biomedical research and the requirements of agricultural deployment. Accordingly, antifungal peptide design for crop protection should not rely on a direct transfer of clinical models, but instead on frameworks specifically trained and validated under agro-relevant biological and environmental conditions.

Future developments should therefore focus on the establishment of curated benchmark datasets and robust predictive pipelines explicitly designed for crop-protection purposes. In particular, next-generation QSAR and ML strategies should move toward multi-objective optimization frameworks that simultaneously integrate antifungal activity with key functional and deployment-related properties. These should include, among others, activity against relevant phytopathogens such as *Botrytis cinerea*, *Fusarium* spp., *Magnaporthe oryzae*, *Alternaria* spp., and *Podosphaera xanthii*, as well as coverage across diverse fungal taxa with distinct infection strategies.

Equally important are environmental and formulation-related constraints that ultimately determine field performance. These include protease resistance, UV stability, pH tolerance, and sustained activity under fluctuating plant-associated conditions, together with solubility, formulation compatibility, leaf-surface retention, rainfastness, and interactions with the plant cuticle. In addition, safety considerations such

as low phytotoxicity and minimal impact on beneficial plant-associated microbiota must be incorporated as explicit optimization criteria. Finally, practical feasibility in terms of synthesis routes, recombinant expression, and scalable production processes should also be considered to ensure translational relevance.

QSAR and ML methodologies already constitute a highly promising and powerful framework for accelerating the discovery and development of antifungal peptides in crop protection. Their value lies in providing a rational, data-driven prioritization strategy whose predictive performance is expected to substantially improve as models are trained and validated with domain-specific, agriculturally relevant datasets, descriptors, and standardized evaluation schemes. Realizing this potential will require a stronger convergence between computational peptide design, plant pathology, formulation science, and experimental validation under real crop-production conditions.

Importantly, rather than representing a limitation, the current state of the field should be regarded as a significant opportunity for innovation and technological expansion. QSAR and ML approaches offer a unique and still underexploited route for translating antifungal peptide discovery into next-generation biofungicide development. In this sense, these methodologies provide a robust conceptual and computational foundation that, once adapted to agricultural endpoints, could enable efficient exploration of peptide chemical space, rational prioritization of candidates, and significantly accelerated development pipelines for sustainable crop protection.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors utilized ResearchRabbit as an additional tool to assist in the collection of relevant literature. ChatGPT was used to help improve the English writing. ChatGPT was also used to support the graphical organization and layout of Fig. 2, whereas the scientific content, terminology, figure labels, final graphical editing, and validation of the information were performed manually by the authors. Following the use of these tools and services, the authors reviewed and edited the content as necessary and take full responsibility for the content of the published article.

## Funding

The publication is part of the project PID2022-136240OB-C21 and PID2022-136240OB-C22, funded by MCIN/AEI/10.13039/501100011033/and by FEDER. A way of making Europe, EU.

## CRediT authorship contribution statement

**Riccardo Zanni:** Conceptualization, Data curation, Methodology, Writing – original draft. **Maria Galvez-Llompart:** Conceptualization, Methodology, Writing – review & editing. **Ignacio Bueso-Bordils:** Data curation. **Alejandro Perez-Garcia:** Project administration, Resources, Supervision. **Dolores Fernandez-Ortuño:** Project administration, Resources. **Facundo Perez-Giménez:** Project administration, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

All data is already available in the manuscript.

## References

[1] Food and Agriculture Organization of the United Nations, The State of Food Insecurity in the World 2013: the Multiple Dimensions of Food Security, FAO, 2013. http://www.fao.org/publications/sofi/2013/en/.
[2] S. Savary, L. Willocquet, S.J. Pethybridge, P. Esker, N. McRoberts, A. Nelson, The global burden of pathogens and pests on major food crops, Nat. Ecol. Evol. 3 (3) (2019) 430–439, https://doi.org/10.1038/s41559-018-0793-y.
[3] Y. Gai, H. Wang, Plant disease: a growing threat to global food security, Agronomy 14 (8) (2024) 1615, https://doi.org/10.3390/agronomy14081615.
[4] United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects 2024: Summary of Results (UN DESA/POP/2024/TR/No. 9), United Nations, 2024.
[5] J. Zhou, X. Zhang, Z. Qu, C. Zhang, F. Wang, T. Gao, Y. Yao, J. Liang, Progress in research on prevention and control of crop fungal diseases in the context of climate change, Agriculture 14 (2024) 1108, https://doi.org/10.3390/agriculture14071108.
[6] Food and Agriculture Organization of the United Nations, The Future of Food and Agriculture – Drivers and Triggers for Transformation (The Future of Food and Agriculture No. 3), FAO, 2022.
[7] M. Figueroa, K.E. Hammond-Kosack, P.S. Solomon, A review of wheat diseases—A field perspective, Mol. Plant Pathol. 19 (6) (2018) 1523–1536, https://doi.org/10.1111/mpp.12618.
[8] C.-M. Gan, T. Tang, Z.-Y. Zhang, M. Li, X.-Q. Zhao, S.-Y. Li, Y.-W. Yan, M.-X. Chen, X. Zhou, Unraveling the intricacies of powdery mildew: insights into colonization, plant defense mechanisms, and future strategies, Int. J. Mol. Sci. 26 (2025) 3513, https://doi.org/10.3390/ijms26083513.
[9] A. Vielba-Fernández, A. Polonio, L. Ruiz-Jiménez, A. de Vicente, A. Pérez-García, D. Fernández-Ortuño, Resistance to the SDH fungicides boscalid and fluopyram in Podosphaera xanthii populations from commercial cucurbit fields in Spain, J. Fungi 7 (9) (2021) 733, https://doi.org/10.3390/jof7090733.
[10] S.A. Alzohairy, L. Heger, N. Nikzainalalam, T.D. Miles, Cross-resistance of succinate dehydrogenase inhibitors (SDHI) in Botrytis cinerea and development of molecular diagnostic tools for SDHI resistance detection, Phytopathology 113 (6) (2023) 998–1009, https://doi.org/10.1094/PHYTO-09-22-0346-R.
[11] Y. Yin, J. Miao, W. Shao, X. Liu, Y. Zhao, Z. Ma, Fungicide resistance: progress in understanding mechanism, monitoring, and management, Phytopathology 113 (2023) 707–718, https://doi.org/10.1094/PHYTO-10-22-0370-KD.
[12] European Food Safety Authority (EFSA), National summary reports on pesticide residue analyses performed in 2023, EFSA Supporting Publication EN-9320 22 (5) (2025) 243, https://doi.org/10.2903/sp.efsa.2025.EN-9320.
[13] Y.M. Zhang, D.X. Ye, Y. Liu, X.Y. Zhang, Y.L. Zhou, L. Zhang, X.L. Yang, Peptides, new tools for plant protection in eco-agriculture, Adv. Agrochem 2 (1) (2023) 58–78, https://doi.org/10.1016/j.aac.2023.01.003.
[14] S. Pan, D. Agyei, J. Jeevanandam, M.K. Danquah, Bio-active peptides: role in plant growth and defense. En Natural bio-active Compounds: Volume 3: Biotechnology, Bioengineering, and Molecular Approaches, Springer, 2019, pp. 1–29, https://doi.org/10.1007/978-981-13-7438-8_1.
[15] D. Sun, Z. Jia, J. Zhu, J. Liu, Y. Chen, Z. Xu, H. Ma, Antimicrobial peptides and their potential applications in plant protection, Agronomy 15 (2025) 1113, https://doi.org/10.3390/agronomy15051113.
[16] P.A.S. Kawmudhi, S. Chathurika, L. Weerasinghe, Applications of antimicrobial peptides in plant pest and disease control, Discov. Plants 2 (2025) 55, https://doi.org/10.1007/s44372-025-00134-2.
[17] E. Montesinos, Antimicrobial peptides and plant disease control, FEMS Microbiol. Lett. 270 (1) (2007) 1–11, https://doi.org/10.1111/j.1574-6968.2007.00683.x.
[18] N. Goyal, K. Singh, Unveiling recent advances in plant peptides and their impact on agriculture and biotechnology, Nucleus (2025), https://doi.org/10.1007/s13237-025-00574-x.
[19] E. Badosa, M. Planas, L. Feliu, L. Montesinos, A. Bonaterra, E. Montesinos, Synthetic peptides against plant pathogenic bacteria, Microorganisms 10 (9) (2022) 1784, https://doi.org/10.3390/microorganisms10091784.
[20] R. Gu, B. Gao, Z. Wang, M. Yang, J. Dai, Y. Wang, Y. Tian, L. Tao, Plant-derived antimicrobial peptides: mechanism of action, safety and integrated applications, J. Agric. Food Chem. 73 (43) (2025) 27161–27176, https://doi.org/10.1021/acs.jafc.5c06949.
[21] N. Armenova, L. Tsigoriyna, A. Arsov, S. Stefanov, K. Petrov, W. Mu, W. Zhang, P. Petrova, Antifungal biocontrol in sustainable crop protection: microbial lipopeptides, polyketides, and plant-derived agents, J. Fungi (Basel, Switzerland) 12 (1) (2025) 22, https://doi.org/10.3390/jof12010022.
[22] M.C.M. Ceron, R.G. Vitale, S.L. Giudicessi, Recent discoveries of antifungal activity in plant antimicrobial peptides, Int. J. Pept. Res. Therapeut. 31 (2025) 4, https://doi.org/10.1007/s10989-024-10661-3.
[23] M. Tetorya, H. Li, A.T. Djami-Tchatchou, G.W. Buchko, K.J. Czymmek, D.M. Shah, Plant defensin MtDef4-derived antifungal peptide with multiple modes of action and potential as a bio-inspired fungicide, Mol. Plant Pathol. 24 (2023) 896–913, https://doi.org/10.1111/mpp.13336.
[24] V. Kovaleva, I. Bukhteeva, O.Y. Kit, I.V. Nesmelova, Plant defensins from a structural perspective, Int. J. Mol. Sci. 21 (2020) 5307, https://doi.org/10.3390/ijms21155307.
[25] A. Fernández, M.L. Colombo, L.M. Curto, G.E. Gómez, J.M. Delfino, F. Guzmán, L. Bakás, I. Malbrán, S.E. Vairo-Cavalli, Peptides derived from the α-Core and γ-Core regions of a putative Silybum marianum flower Defensin show antifungal activity against Fusarium graminearum, Front. Microbiol. 12 (2021) 632008, https://doi.org/10.3389/fmicb.2021.632008.

[26] T. Li, L. Li, F. Du, L. Sun, J. Shi, M. Long, Z. Chen, Activity and mechanism of action of antifungal peptides from microorganisms: a review, Molecules 26 (2021) 3438, https://doi.org/10.3390/molecules26113438.
[27] X. Tang, Application of antimicrobial peptides in plant protection: making use of the overlooked merits, Front. Plant Sci. 14 (2023) 1139539, https://doi.org/10.3389/fpls.2023.1139539.
[28] H. Li, S.L.S. Velivelli, D.M. Shah, Antifungal potency and modes of action of a novel olive tree defensin against closely related Ascomycete fungal pathogens, Mol. Plant Microbe Interact. 32 (2019) 1649–1664, https://doi.org/10.1094/MPMI-08-19-0224-R.
[29] Z. Lai, X. Yuan, H. Chen, Y. Zhu, N. Dong, A. Shan, Strategies employed in the design of antimicrobial peptides with enhanced proteolytic stability, Biotechnol. Adv. 59 (2022) 107962, https://doi.org/10.1016/j.biotechadv.2022.107962.
[30] O. Al Musaimi, L. Lombardi, D.R. Williams, F. Albericio, Strategies for improving peptide stability and delivery, Pharmaceuticals 15 (10) (2022) 1283, https://doi.org/10.3390/ph15101283.
[31] J. Lin, L. Wen, Y. Zhou, S. Wang, H. Ye, J. Su, J. Li, J. Shu, J. Huang, P. Zhou, PepQSAR: a comprehensive data source and information platform for peptide quantitative structure-activity relationships, Amino Acids 55 (2) (2023) 235–242, https://doi.org/10.1007/s00726-022-03219-4.
[32] D. de Raffele, I.M. Ilie, Unlocking novel therapies: cyclic peptide design for amyloidogenic targets through synergies of experiments, simulations, and machine learning, Chem. Commun. 60 (4) (2024) 632–645, https://doi.org/10.1039/D3CC04630C.
[33] R. Zanni, M. Galvez-Llompart, R. Garcia-Domenech, J. Galvez, What place does molecular topology have in today's drug discovery? Expert Opin. Drug Discov. 15 (10) (2020) 1133–1144, https://doi.org/10.1080/17460441.2020.1770223.
[34] F. Wan, F. Wong, J.J. Collins, C. de la Fuente-Nunez, Machine learning for antimicrobial peptide identification and design, Nat. Rev. Bioeng. 2 (5) (2024) 392–407, https://doi.org/10.1038/s44222-024-00152-x.
[35] T.S. Thind, Changing trends in discovery of new fungicides: a perspective, Indian Phytopathol. 74 (2021) 875–883, https://doi.org/10.1007/s42360-021-00411-6.
[36] T. Joshi, H. Sehgal, S. Puri, Karnika, T. Mahapatra, M. Joshi, P.R. Deepa, P. K. Sharma, ML-based technologies in sustainable agro-food production and beyond: tapping the (semi) arid landscape for bioactives-based product development, J. Agric. Food Res. 18 (2024) 101350, https://doi.org/10.1016/j.jafr.2024.101350.
[37] M. Goles, A. Daza, G. Cabas-Mora, L. Sarmiento-Varón, J. Sepúlveda-Yañez, H. Anvari-Kazemabad, M.D. Davari, R. Uribe-Paredes, Á. Olivera-Nappa, M. A. Navarrete, D. Medina-Ortiz, Peptide-based drug discovery through artificial intelligence: towards an autonomous design of therapeutic peptides, Briefings Bioinf. 25 (4) (2024), https://doi.org/10.1093/bib/bbae275 bbae275.
[38] A. Pandi, D. Adam, A. Zare, V.T. Trinh, S.L. Schaefer, M. Burt, B. Klabunde, E. Bobkova, M. Kushwaha, Y. Foroughijabbari, P. Braun, C. Spahn, C. Preußer, E. Pogge von Strandmann, H.B. Bode, H. von Buttlar, W. Bertrams, A.L. Jung, F. Abendroth, B. Schmeck, T.J. Erb, Cell-free biosynthesis combined with deep learning accelerates de novo development of antimicrobial peptides, Nat. Commun. 14 (1) (2023) 7197, https://doi.org/10.1038/s41467-023-42434-9.
[39] B. Kergaravat, L. Plener, M. Castan, D. Grizard, É. Chabrière, D. Daudé, Enzymes, proteins, and peptides as promising biosolutions for crop production, J. Agric. Food Chem. 73 (19) (2025) 11546–11555, https://doi.org/10.1021/acs.jafc.5c01107.
[40] P. Szymczak, W. Zarzecki, J. Wang, Y. Duan, J. Wang, L.P. Coelho, C. de la Fuente-Nunez, E. Szczurek, AI-Driven antimicrobial peptide discovery: mining and generation, Acc. Chem. Res. 58 (12) (2025) 1831–1846, https://doi.org/10.1021/acs.accounts.0c00594.
[41] J. Yan, J. Cai, B. Zhang, Y. Wang, D.F. Wong, S.W.I. Siu, Recent progress in the discovery and design of antimicrobial peptides using traditional machine learning and deep learning, Antibiotics (Basel, Switzerland) 11 (10) (2022) 1451, https://doi.org/10.3390/antibiotics11101451.
[42] J. Zhang, X. Huang, S. Wang, Y. Chen, Large-scale discovery of antifungal peptides using quantitative structure–activity relationships and virtual screening, ACS Med. Chem. Lett. 13 (2) (2022) 298–304, https://doi.org/10.1021/acsmedchemlett.1c00556.
[43] R. Seiad Ahmadnezhad, M. Shams-Ghahfarokhi, F. Jamzivar, A. Eslamifar, A. Sohrabi, M. Razzaghi-Abyaneh, Next-generation antifungal peptide discovery: the synergy of artificial intelligence and omics technologies, World J. Microbiol. Biotechnol. 41 (12) (2025) 456, https://doi.org/10.1007/s11274-025-04662-7.
[44] F. Wan, D. Kontogiorgos-Heintz, C. de la Fuente-Nunez, Deep generative models for peptide design, Digit. Discov. 1 (3) (2022) 195–208, https://doi.org/10.1039/D1DD00024A.
[45] C.M. Van Oort, J.B. Ferrell, J.M. Remington, S. Wshah, J. Li, AMPGAN v2: machine learning-guided design of antimicrobial peptides, J. Chem. Inf. Model. 61 (5) (2021) 2198–2207, https://doi.org/10.1021/acs.jcim.0c01441.
[46] C. Li, D. Sutherland, A. Richter, L. Coombe, A. Yanai, R.L. Warren, M. Kotkoff, F. Hof, L.M.N. Hoang, C.C. Helbing, I. Birol, De novo synthetic antimicrobial peptide design with a recurrent neural network, Protein Sci. 33 (8) (2024) e5088, https://doi.org/10.1002/pro.5088.
[47] X.F. Wang, J.Y. Tang, J. Sun, S. Dorje, T.Q. Sun, B. Peng, X.W. Ji, Z. Li, X.E. Zhang, D.B. Wang, ProT-Diff: a modularized and efficient strategy for de novo generation of antimicrobial peptide sequences by integrating protein language and diffusion models, Adv. Sci. 11 (43) (2024) e2406305, https://doi.org/10.1002/advs.202406305.
[48] M.A. Zervou, E. Doutsi, Y. Pantazis, P. Tsakalides, De Novo Antimicrobial Peptide Design with Feedback Generative Adversarial Networks, Int. J. Mol. Sci. 25 (10) (2024) 5506, https://doi.org/10.3390/ijms25105506.
[49] D.S. Berman, L.M. Lewis, T.D. Curtis, O.N. Tiburzi, D.F.Q. Smith, A. Casadevall, L. J. Dunphy, Fung-AI: an AI/ML-driven pipeline for antifungal peptide discovery, PLoS Comput. Biol. 22 (6) (2026) e1014105, https://doi.org/10.1371/journal.pcbi.1014105.
[50] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57 (12) (2014) 4977–5010, https://doi.org/10.1021/jm4004285.
[51] C. Rodríguez-Cerdeira, A. Molares-Vila, C.D. Sánchez-Cárdenas, J.S. Velásquez-Bámaca, E. Martínez-Herrera, Bioinformatics approaches applied to the discovery of antifungal peptides, Antibiotics 12 (3) (2023) 566, https://doi.org/10.3390/antibiotics12030566.
[52] C. Chen, Y. Tan, Y. Liu, M. Sun, Z. Wang, X. Zheng, Z. Shen, M. Zhao, Applications of machine learning-based quantitative structure activity relationship (ML-QSAR) models in environmental science, Environ. Res. 294 (2026) 123853, https://doi.org/10.1016/j.envres.2026.123853.
[53] S. Bouamrane, A. Khaldan, M. Alaqarbeh, A. Sbai, M.A. Ajana, T. Lakhlifi, M. Bouachrine, H. Maghat, Computational integration for antifungal 1,2,4-triazole inhibitors design: QSAR, molecular docking, molecular dynamics simulations, ADME/Tox, and retrosynthesis studies, Chem. Phys. Impact 8 (2024) 100502, https://doi.org/10.1016/j.chphi.2024.100502.
[54] M.R. Borkar, R.R.S. Pissurlenkar, E.C. Coutinho, Mapping activity elements of protegrin antimicrobial peptides by HomoSAR, RSC Adv. 5 (96) (2015) 78790–78798, https://doi.org/10.1039/C5RA14402G.
[55] S. Yousefinejad, M. Bagheri, A.A. Moosavi-Movahedi, Quantitative sequence-activity modeling of antimicrobial hexapeptides using a segmented principal component strategy: an approach to describe and predict activities of peptide drugs containing L/D and unnatural residues, Amino Acids 47 (1) (2015) 125–134, https://doi.org/10.1007/s00726-014-1850-8.
[56] S.A. Pinacho-Castellanos, C.R. García-Jacas, M.K. Gilson, C.A. Brizuela, Alignment-free antimicrobial peptide predictors: improving performance by a thorough analysis of the largest available data set, J. Chem. Inf. Model. 61 (6) (2021) 3141–3157, https://doi.org/10.1021/acs.jcim.1c00251.
[57] D.H. Chang, M.R. Lee, N. Wang, D.M. Lynn, S.P. Palecek, Establishing quantifiable guidelines for antimicrobial α/β-peptide design: a partial least-squares approach to improve antimicrobial activity and reduce mammalian cell toxicity, ACS Infect. Dis. 9 (12) (2023) 2632–2651, https://doi.org/10.1021/acsinfecdis.3c00468.
[58] F. Lobo, M.S. González, A. Boto, J.M. Pérez de la Lastra, Prediction of antifungal activity of antimicrobial peptides by transfer learning from protein pretrained models, Int. J. Mol. Sci. 24 (12) (2023) 10270, https://doi.org/10.3390/ijms241210270.
[59] A. Medvedeva, C. Vasnetsov, V. Vasnetsov, A.B. Kolomeisky, Antimicrobial peptides as broad-spectrum therapeutics: computational analysis to identify universal physicochemical features responsible for multitarget activity, J. Phys. Chem. Lett. 15 (50) (2024) 12416–12424, https://doi.org/10.1021/acs.jpclett.4c03197.
[60] K. Yin, R. Li, S. Zhang, Y. Sun, L. Huang, M. Jiang, D. Xu, W. Xu, Deep learning combined with quantitative structure–activity relationship accelerates de novo design of antifungal peptides, Adv. Sci. 12 (13) (2025) e2412488, https://doi.org/10.1002/advs.202412488.
[61] P. Agrawal, S. Bhalla, K. Chaudhary, R. Kumar, M. Sharma, G.P.S. Raghava, In silico approach for prediction of antifungal peptides, Front. Microbiol. 9 (2018) 323, https://doi.org/10.3389/fmicb.2018.00323.
[62] A. Tyagi, R. Kumar, S. Singh, K. Chaudhary, A. Gautam, G.P.S. Raghava, PhytoAFP: in silico approaches for designing plant-derived antifungal peptides, Antibiotics 10 (6) (2021) 652, https://doi.org/10.3390/antibiotics10060652.
[63] V. Frecer, QSAR analysis of antimicrobial and haemolytic effects of cyclic cationic antimicrobial peptides derived from protegrin-1, Bioorg. Med. Chem. 14 (2) (2006) 606–616, https://doi.org/10.1016/j.bmc.2005.08.056.
[64] H. Wang, L. Yao, J. Chen, Z. Ding, X. Ou, C. Zhang, J. Zhao, Y. Han, Antifungal peptide P852 effectively controls Fusarium oxysporum, a wilt-causing fungus, by affecting glucose metabolism and amino acid metabolism as well as damaging mitochondrial function, J. Agric. Food Chem. 71 (2023) 19638–19651, https://doi.org/10.1021/acs.jafc.3c07953.