

Review

# Bioactive Molecules from the Innate Immunity of Ascidians and Innovative Methods of Drug Discovery: A Computational Approach Based on Artificial Intelligence

Laura La Paglia <sup>1</sup>, Mirella Vazzana <sup>2</sup>, Manuela Mauro <sup>2</sup>, Alfonso Urso <sup>1</sup>, Vincenzo Arizza <sup>2</sup>  
and Aiti Vizzini <sup>2,\*</sup>

<sup>1</sup> Istituto di Calcolo e Reti ad Alte Prestazioni–Consiglio Nazionale delle Ricerche, Via Ugo La Malfa 153, 90146 Palermo, Italy; laura.lapaglia@icar.cnr.it (L.L.P.); alfonso.urso@icar.cnr.it (A.U.)

<sup>2</sup> Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche–Università di Palermo, Via Archirafi 18, 90100 Palermo, Italy; mirella.vazzana@unipa.it (M.V.); manuela.mauro01@unipa.it (M.M.); vincenzo.arizza@unipa.it (V.A.)

\* Correspondence: aiti.vizzini@unipa.it

**Abstract:** The study of bioactive molecules of marine origin has created an important bridge between biological knowledge and its applications in biotechnology and biomedicine. Current studies in different research fields, such as biomedicine, aim to discover marine molecules characterized by biological activities that can be used to produce potential drugs for human use. In recent decades, increasing attention has been paid to a particular group of marine invertebrates, the Ascidians, as they are a source of bioactive products. We describe omics data and computational methods relevant to identifying the mechanisms and processes of innate immunity underlying the biosynthesis of bioactive molecules, focusing on innovative computational approaches based on Artificial Intelligence. Since there is increasing attention on finding new solutions for a sustainable supply of bioactive compounds, we propose that a possible improvement in the biodiscovery pipeline might also come from the study and utilization of marine invertebrates' innate immunity.

**Keywords:** bioactive molecules; marine invertebrate; artificial intelligence; innate immunity



**Citation:** La Paglia, L.; Vazzana, M.; Mauro, M.; Urso, A.; Arizza, V.; Vizzini, A. Bioactive Molecules from the Innate Immunity of Ascidians and Innovative Methods of Drug Discovery: A Computational Approach Based on Artificial Intelligence. *Mar. Drugs* **2024**, *22*, 6. <https://doi.org/10.3390/md22010006>

Academic Editor: Francesc Xavier Avilés

Received: 20 November 2023

Revised: 12 December 2023

Accepted: 17 December 2023

Published: 20 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ascidians are tunicates, marine invertebrate chordates, considered the sister group of vertebrates [1–4]. They owe their name to the tunic, an epidermally secreted external layer that envelops the body. The tunic is composed of an ECM rich in collagen and tunicin (a form of cellulose) and also rich in immune cells [5–9]. The tunic also contains proteins with 3,4-dihydroxyphenylalanine (DOPA), with a catechol moiety involved in the first line of immune reaction [10] and wound healing [11,12], and 3,4,5-trihydroxyphenylalanine (TOPA) with a pyrogallol moiety [13]. Ascidians are the source of many bioactive molecules belonging to a wide variety of chemical categories [14,15] and with potential health applications, including cytotoxic, antimetabolic, antiviral and antimicrobial compounds [16–19]. Most of the metabolites synthesized by ascidians contribute to creating the physico-chemical barrier preventing the entrance of foreign organisms into the internal fluids or the colonization of the tunic by encrusting organisms. The most represented chemical classes among the bioactive secondary metabolites isolated from tunicates are alkaloids, polyketides, and peptides [18]. Cytotoxicity against mammalian cell lines and anti-proliferative activity were the most frequently assigned bioactivities. Compounds with cytotoxic and antineoplastic properties isolated from ascidians belong to disparate chemical classes, and three of them have entered clinical trials [20]. Due to their key phylogenetic position in chordate evolution, the ascidians are a powerful model for studying innate immunity [4]. They possess an exclusively innate immune system, including inflammatory, humoral, and cellular

responses. On an evolutionary level, inflammation is a highly conserved phenomenon and appears to be an essential first line of defense for both invertebrates and vertebrates. The innate immune system is the major contributor to acute inflammation [21,22], a rapid protective response to microbial infection, tissue injury, and insults [23], and the principal promoter of inflammatory responses often involves infection by microbial invaders or exposure to foreign particles/irritants/pollutants [24].

When host cells capable of innate immune responses encounter pathogenic microbes or other foreign or host irritants, the inflammatory response initiates within minutes. The host cells recognize the stimulus through various sensing mechanisms involving transmembrane receptors. In *Ciona robusta* (previously *Ciona intestinalis*), these interactions transmit signals to the nucleus, resulting in the activation and regulation of numerous genes via both transcriptional and post-transcriptional mechanisms [21,25–33], such as antimicrobial peptides and complement factors [34,35] and proinflammatory cytokines and chemokines which activate endothelial cells and recruit immune system cells to the infection site [36–38].

The immune system is made up of a complex and dynamic network of cell subsets and mediators that promote host defense from infectious agents or tumor cells and maintain immunological tolerance in the organisms [39,40]. Vertebrate immunity is classically divided into innate and adaptive immune systems that act interdependently based on bidirectional crosstalk [41]. It is well known that the innate immune system provides the first line of defense in immune response and can induce and regulate many different adaptive immunity functions [42].

This review provides an overview of the bioactive molecules isolated from the innate immune system response of ascidians and, at the same time, describes a new approach combining omics technologies and new bioinformatic strategies for drug discovery based on Artificial Intelligence to reduce costly and time-consuming conventional laboratory testing, validation, and synthetic procedures and accelerate the drug discovery process. We also propose the study of omics data obtained from the innate immune processes of ascidians as an excellent source of innovative bioactive molecules such as antimicrobial, anticancer, and immunomodulatory peptides.

## 2. Bioactive Molecules in Ascidians

Bioactive peptides (BPs) are composed of protein fragments or peptides with beneficial metabolic and physiological functions that promote human health; thus, they are excellent molecules for studying human health and disease and potential therapeutics [43,44]. Most have similar structures, such as <20 amino acid residue lengths, and contain hydrophobic residues [45,46]. Based on their modes of action, different classes of BP are evidenced: anti-cancer (ACP), antiviral, antimicrobial (AMP), anti-oxidative stress, and immunomodulatory peptides [15–19,47,48].

In the following sections, only AMP and immunomodulatory peptides will be discussed, as they are the only bioactive molecules shown in ascidians.

### 2.1. Bioactive Molecules with Antimicrobial Activity

AMPs are a class of small peptides that exist widely in nature, and they are an important part of the innate immune system of different organisms. AMPs have a broad range of inhibitory effects against bacteria, fungi, parasites, and viruses. The emergence of antibiotic-resistant microorganisms and increasing concerns about the use of antibiotics have resulted in the development of AMPs, which have good application prospects in medicine, food, animal husbandry, agriculture, and aquaculture. Microorganism resistance to antimicrobials is becoming increasingly severe with the abuse of antibiotics in medicine, agriculture, and animal husbandry. The prevalence of vancomycin-resistant enterococcus (VRE) and methicillin-resistant *Staphylococcus aureus* (MRSA) is increasing in clinical medicine, so countermeasures are urgently needed to address these bacterial infections. Research on AMPs is continuously developing, and AMP databases store a considerable amount of data

on AMPs. A massive variety of antimicrobials has been extracted from tunicates. They belong to disparate chemical classes, such as polysulfides, alkyl sulfates, terpenes, amino alcohols, spiroketals, alkaloids, furanones, peptides, and others [15]. Some of them are synthesized by symbiotic organisms colonizing the tunic or the internal fluids [49,50].

Most of the known AMPs are produced by ascidian-circulating cells, mainly immunocytes (i.e., cells involved in immune responses) for defense purposes [51–60] (Table 1).

**Table 1.** Bioactive molecules in ascidians from immunocytes: antimicrobial and immunomodulator peptides.

Species	Peptide	Activity	References
<i>Halocynthia roretzi</i>	Halocyamines A and B	Antimicrobial	[51]
<i>Halocynthia aurantium</i>	Dicynthaurin halocidin	Antimicrobial	[54]
<i>Halocynthia papillosa</i>	Halocintin and papillosin	Antimicrobial	[60]
<i>Styela plicata</i>	Clavanins A-D	Antimicrobial	[55]
<i>Ciona intestinalis</i>	Antimicrobial peptides	Antimicrobial	[34,52,61]
<i>Ciona robusta</i>	C8, CrCp	Immunomodulatory	[62,63]

In *Halocynthia roretzi*, the tetrapeptides halocyamines A and B are produced by cytotoxic morula cells (MCs) [51], and their cytotoxic activity is likely related to their diphenol rings, representing suitable substrates for the enzyme phenoloxidase (PO), which is also stored inside MCs. The enzyme induces oxidative stress by oxidizing phenols to quinones with the consequent production of ROS [64]. The hemocytes of species *H. aurantium* synthesize the peptide dicynthaurin and the cationic peptide halocidin [54]. The native peptide of halocidin has a mass of 3443 Da and comprises two different subunits containing 18 amino acid residues and 15 residues which are linked covalently by a single cystine disulfide bond. Two different monomers were separately synthesized to make three additional isoforms (15-residue homodimer, 18-residue homodimer, heterodimer). Antimicrobial assays performed with synthetic peptides of halocidin confirmed that congeners of the 18-residue monomer were more active than those of the 15-residue monomer MRSA and multidrug-resistant *Pseudomonas aeruginosa*.

Hemocytes from the solitary tunicate *Styela clava* contained a family of four  $\alpha$ -helical antimicrobial peptides that were purified, sequenced, and named clavanins A, B, C, and D. Clavanins A–D (histidine-rich,  $\alpha$ -helix peptides) [49] and clavaspirin are synthesized by *Styela clava* MCs [56]. In lysates of hemocytes of the same species, five cationic antimicrobial peptides, called styelins, were identified and isolated [57,58]. In hemocytes of *Styela plicata*, the octapeptide plicatamide was isolated [59]. In the tunicates *Microcosmus sabatieri* and *Halocynthia papillosa*, antimicrobial activities were detected, and two novel peptides, halocintin and papillosin, were isolated and characterized. These molecules display antibacterial activity against Gram-positive and Gram-negative bacteria. A combination of Edman degradation and mass spectrometry obtained a complete peptide characterization. The mature molecules of halocintin and papillosin comprise 26 and 34 amino acid residues, respectively [60].

The enormous quantity of genomic data has become a promising source of putative AMPs due to progress in bioinformatics [65–68]. In *C. intestinalis*, using genome and expressed sequence tag (EST) data, a putative gene family has been identified exhibiting several structural features typical of AMPs. The synthetic peptide exerted potent antimicrobial activity against various bacteria and against the yeast *Candida albicans*, but it was not cytolytic for mammalian erythrocytes. Using the synthetic peptide as an antigen, specific antibodies were generated, and the natural parent molecule was localized to a compartment of a distinct hemocyte type, the univacuolar refractile granulocytes [53]. Furthermore, a gene family coding for putative AMPs was identified in the EST database of *C. intestinalis* and subsequently identified and cloned from the Northern European *Ciona* subspecies. Molecular analysis revealed that the natural peptide is synthesized and stored in a distinct hemocyte type, the univacuolar non-refractile granulocytes, and that the expression of the

gene is markedly upregulated in hemocytes after immune challenge. The peptide proved highly effective against Gram-negative and Gram-positive bacteria, including several human and marine pathogens, as well as the yeast *C. albicans*. Using two different methods, it was demonstrated that the peptide kills Gram-negative and Gram-positive bacteria by permeabilizing their cytoplasmic membranes. Circular Dichroism (CD) spectroscopy revealed that in the presence of liposomes composed of negatively charged phospholipids, the peptide undergoes a conformational change and adopts an alpha-helical structure. Moreover, the peptide was virtually non-cytolytic for mammalian erythrocytes. Hence, this designed AMP may represent a valuable template for developing novel antibiotics [52].

Ci-MAM-A24, a synthetic AMP derived from a peptide precursor from immune cells of *C. intestinalis*, is potentially active against representatives of Gram-positive and Gram-negative bacteria by permeabilizing their cytoplasmic membrane. The activity of Ci-MAM-A24 against different bacterial pathogens which frequently cause therapeutic problems was tested. Fedders et al. tested the killing capacity of Ci-MAM-A24 against clinically important anaerobic bacteria as well as multiresistant aerobic strains such as MRSA, VRE, extended-spectrum  $\alpha$ -lactamase-producers, and multiple-resistant *Pseudomonas aeruginosa*, and all strains proved to be highly susceptible to Ci-MAM-A24 at low concentrations [69].

Furthermore, an in silico screening method has been developed based on further criteria such as size, amphipathicity, and aggregation propensity, by which 22 potential LCAMP candidates in the *Ciona* genome were computationally predicted. Among these LCAMP candidates, five novel salt-resistant LCAMPs with broad-spectrum antimicrobial activity were experimentally confirmed. This strategy was also successfully applied to the *Xenopus tropicalis* genome, suggesting that this method could apply to the in silico screening of any genome [34].

Finally, Lu et al. [61] investigated the potential sORFs encoding AMPs in *C. intestinalis*, and over 180 peptides deduced from the sORFs were predicted to be AMPs. Among the ten peptides tested, six were found to have significant EST matches, providing strong evidence for gene expression; five were proved to be active against the bacterial strains.

## 2.2. Bioactive Molecules with Immunomodulatory Effects

A recent idea is to use invertebrates as a source of molecules with potential immunoregulatory activities to improve strategies for modulating human immune system responses [65,66]. The innate immune system is composed of many interdependent cell types and mediators. It is one of the most critical natural systems for protection against many harmful bacteria, viruses, parasites, and fungi in human health, and against autoimmune diseases, cancer, allergies, and infections [70,71].

Preliminary studies have shown evidence supporting a complex interaction between the immune system and tumors [72]. Several innate system immunomodulators have been identified; these include cytokines (interleukins, interferons, and chemokines), substances isolated from microorganisms and fungi (lipopolysaccharides; LPS), and substances isolated from plants (polysaccharides and phenolic compounds) [73]. Tumor cells secrete altered protein products that must be recognized as foreign by the immune effector cells such as B, T, natural killer and natural killer T cells, and type I and II interferons, and perforin which are able to destroy tumor cells [74,75]. Therefore, the enhancement of the host immune response is one of the most important methods for inhibiting tumor growth and maintaining cellular homeostasis without harming the host.

The selective modulation of immunity is an emerging concept driven by the tremendous advances in our understanding of this crucial host defense system. Invertebrates have drawn researchers' interest as potential sources of new bioactive molecules owing to their immunomodulatory activities. An LPS challenge in the ascidian *C. intestinalis* generates the transcript, Ci8 short, with cis-regulatory elements in the 3' UTR region which is essential for shaping innate immune responses. The derived amino acidic sequence from in silico analysis showed specific binding to human major histocompatibility complex (MHC) class I and class II alleles. The role of Ci8 short peptide (Table 1) was investigated in a more

evolved immune system using human peripheral blood mononuclear cells (PBMCs) as an in vitro model. The biological activities of this molecule include the activation of the 70 kDa TCR  $\zeta$  chain associated protein kinase (ZAP-70) and T cell receptor (TCR) V $\beta$  oligo clonal selection on CD4+ T lymphocytes as well as increased proliferation and IFN- $\gamma$  secretion. Furthermore, Ci8 short affects CD4+/CD25high-induced regulatory T cells (iTreg) subset selection, which co-expressed the functional markers TGF- $\beta$ 1/latency-associated protein (LAP) and CD39/CD73 [62].

Furthermore, Colombo et al. [63] evaluated the 3D structure of the C8 short-derived *C. robusta* chemo-attractive peptide (CrCP) (Table 1) by homology modeling, which showed that CrCP displayed structural characteristics already reported for a short domain of the vertebrate CRK gene, suggesting its possible involvement in cell migration mechanisms. The biological activity of CrCP was studied in vitro using a primary human dermal cell line. In vitro assays demonstrated that CrCP could induce the motility of HuDe cells in both wound healing and chemo-attractive experiments. Furthermore, CrCP modulates the expression of the matrix metalloproteinase-7 (MMP-7) and E-cadherin genes, and it induces the activation of the NF- $\kappa$ B signaling pathway.

### 3. Bioactive Molecule Identification through Omics Technologies

Classical methods for peptide analysis have relied principally on targeted immunoassays which enable the biochemical purification of bioactive peptides from tissues or cells by identifying fractions with a desired bioactivity [76]. By performing multiple rounds of purification, bioactive peptides could be subsequently identified. This approach discovered many peptides in ascidians, including the isolation of Clavanins in *Styela clava* hemocytes [55] and Halocyanine in the ascidian *Halocynthia roretzi* [51]. During the last decade, technological innovations and the advent of omics data have led to an explosion of biological information. Indeed, different strategies, including transcriptomics and proteomics approaches, such as next-generation sequencing or mass spectrometry, have substituted the classical methods for peptide analysis, producing a considerable amount of biological data. The increasing amount of biological data produced was accompanied by the use of bioinformatics tools to support big data analysis.

Fedders et al.'s initial efforts to integrate genomic data and bioinformatics [53] proceeded through a reverse search for AMPs in *C. robusta*. They used the completed genome project and the substantial amount of EST data available as a screening matrix in association with bioinformatics techniques for the design of synthetic AMPs with in silico tools. Another recent approach to investigating bioactive peptides in *C. robusta* by integrating omics data analysis with bioinformatics focused on the study of the 3D structure of the C8short-derived CrCP, which was evaluated by homology modeling. In vitro studies using a primary human dermal cell line (HuDe) evaluated the biological activity of CrCP. A short domain of the vertebrate CRK gene was identified, suggesting its possible involvement in cell migration mechanisms [75].

Later, Kawada et al. [77] showed how omics studies associated with AI algorithms could contribute to the elucidation of gene expression profiles. These revealed key regulatory genes for *Ciona* follicle growth, maturation, and ovulation, verifying essential and novel molecular mechanisms underlying these biological events with the contribution of machine learning techniques. Furthermore, AI has been employed to solve some of the most challenging issues of bioinformatics, including protein structure prediction, homology searches, multiple alignment and phylogeny construction, genomic sequence analysis, gene finding, and more. Thus, combining omics data with AI and single-cell technologies, Kawada et al. [77] paved the way for investigating in greater detail the nervous, neuroendocrine, and endocrine systems of ascidians and the molecular and functional evolution and diversity of peptidergic regulatory networks throughout chordates. Franchi et al. [78] reported the identification, by mining the *B. schlosseri* transcriptome, of a transcript for a putative styelin-like AMP named botryllin, which is actively transcribed by morula cells

(MCs). The synthetic peptide, obtained from *in silico* translation of the transcript, exerted toxic activity toward bacterial and unicellular yeast cells.

Additional omics data may come from proteome analyses and can also be useful for identifying proteins and peptides with molecule bioactive features, such as AMPs [79]. One of the most widespread proteomic approaches is liquid chromatography coupled with mass spectrometry (LC-MS) [80]. The analysis of spectra obtained by mass spectrometry consists of protein profiling, peptide mapping and identification, and protein quantification. Matos et al. [81], through shotgun proteomics of the ascidian tunic, provided new insights on host–microbe interactions by revealing diverse AMPs. They reported different proteins associated with immune mechanisms of invertebrates, as in the case of the metazoan Down syndrome cell adhesion molecule-like protein, which was detected in *Molgula* sp. samples and related to immune mechanisms mediating phagocytosis and adherence of bacteria [82], or the barrier to autointegration factor (BAF), involved in innate immune response as an inhibitor of exogenous viral DNA replication and involved in host defense response [83], revealing the tunic as a very active tissue in terms of bioactive compound production [81].

A recent study on innate immune response mechanisms to microbial stimuli in a *C. robusta* invertebrate model showed the importance of multi-omics analysis and bioinformatics approaches. Indeed, the intersection of large-scale sequencing or other “-omics” approaches, as the combination of transcriptomic and proteomic data analysis, allows the better investigation of the cellular pathways and biological processes affected by microbial treatments and the investigation of host responses to PAMPs in different physiological conditions and at various stages of maturation of the immune system, thus filling potential gaps concerning expression differences observed in the synthesis of proteins related to mRNA expression [84].

#### 4. AI-Based Computational Approaches and Their Role in Drug Discovery

To guide the reader in the following methodological sections, we will briefly introduce the main concepts of AI and its principal subclasses, and then we will focus on AI approaches applied to drug discovery.

AI is a technology-based system that can mimic human intelligence through features such as reasoning, knowledge representation, and solution research. At the same time, it does not threaten to completely replace human physical presence. Moreover, AI can help manage the massive amount of data produced and give decision-making support in clinical and translational research.

Machine learning (ML) is a subclass of AI. Through ML processes, computers can learn without explicitly being programmed. It performs prediction and classification tasks through pattern detection without using defined rules [85]. There are two ML algorithm classes: supervised and unsupervised learning.

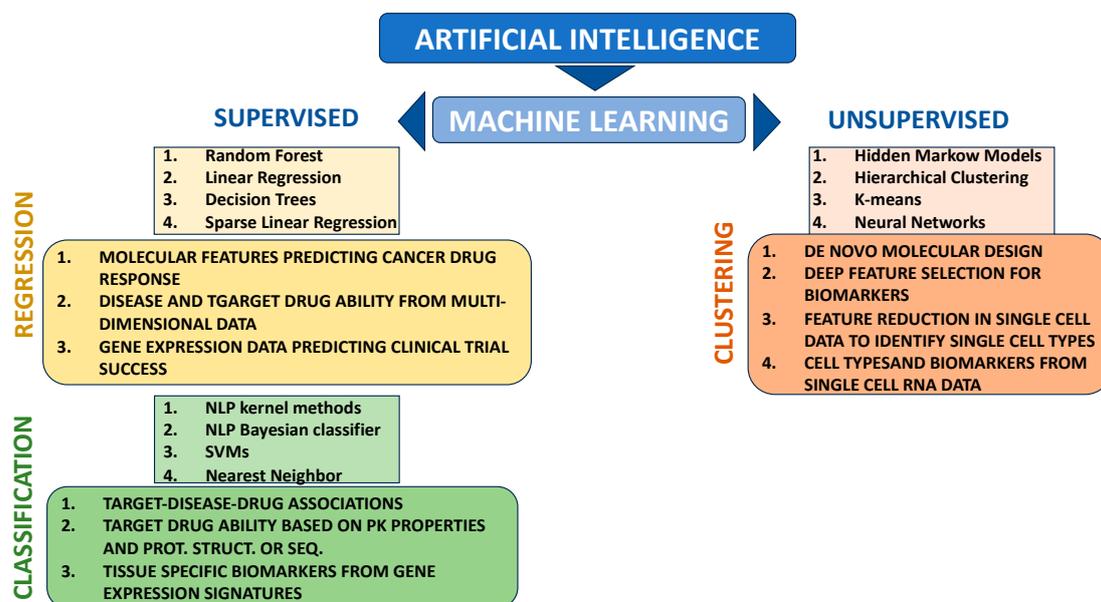
In supervised learning, the datasets are designed to train or “supervise” algorithms to classify data or to accurately predict outcomes. Moreover, supervised learning problems can further be divided into “classification” and “regression” tasks: classification assigns test data to specific categories, and regression can predict a continuous numerical output, helping in establishing a relationship among the variables by estimating how one variable affects the other. Another subclass of ML is classification, which is “unsupervised” as it takes references from datasets consisting of input data without labeled responses [85–92].

Due to the massive amount of data produced by NGS techniques, AI has made a significant contribution to data analysis. Also, in bioactive compound analysis, ML algorithms are the basis of many prediction and analysis methods. Deep Learning (DL) algorithms are a subset of ML [93–98].

Drug discovery is the process through which new medications against diseases are discovered. It consists of a combination of a wide variety of technologies and expertise aimed at finding potential drugs against specific targets. Typical examples of drug discovery tasks are drug-target prediction [99], bioavailability prediction [100], and *de novo* drug design [101]. Moreover, there is also the main category of pharmaceutical analysis that

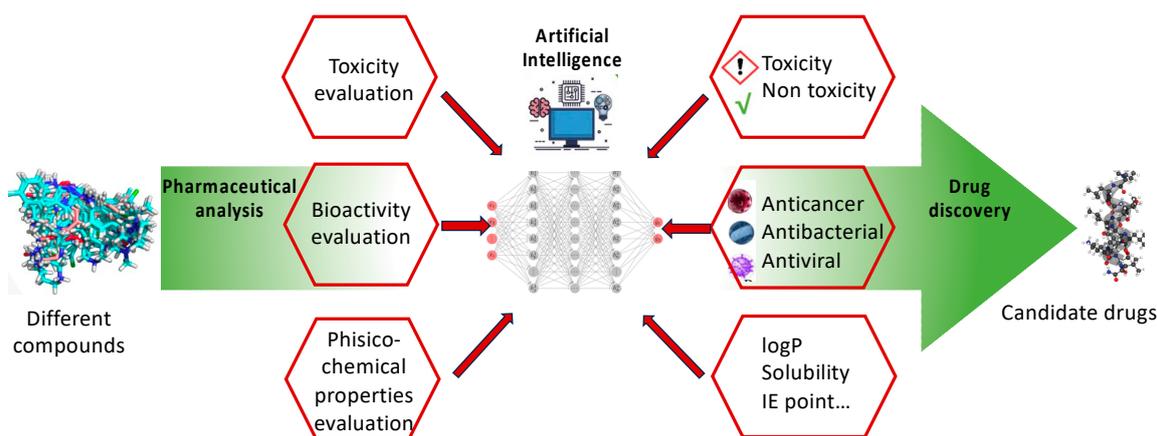
groups these examples of drug discovery, involving different subtasks such as toxicity analysis, bioactivity evaluation, and physicochemical property analysis.

Advancements in AI techniques have revolutionized their applications to this field of research (Figure 1). Indeed, many AI approaches have been developed during the last decade specifically for this aim, allowing the acceleration of the drug discovery process and, at the same time, reducing the high-cost characteristics of conventional methods in terms of money and time [102].



**Figure 1.** Principal classes and subclasses of AI algorithms and their involvement in drug discovery.

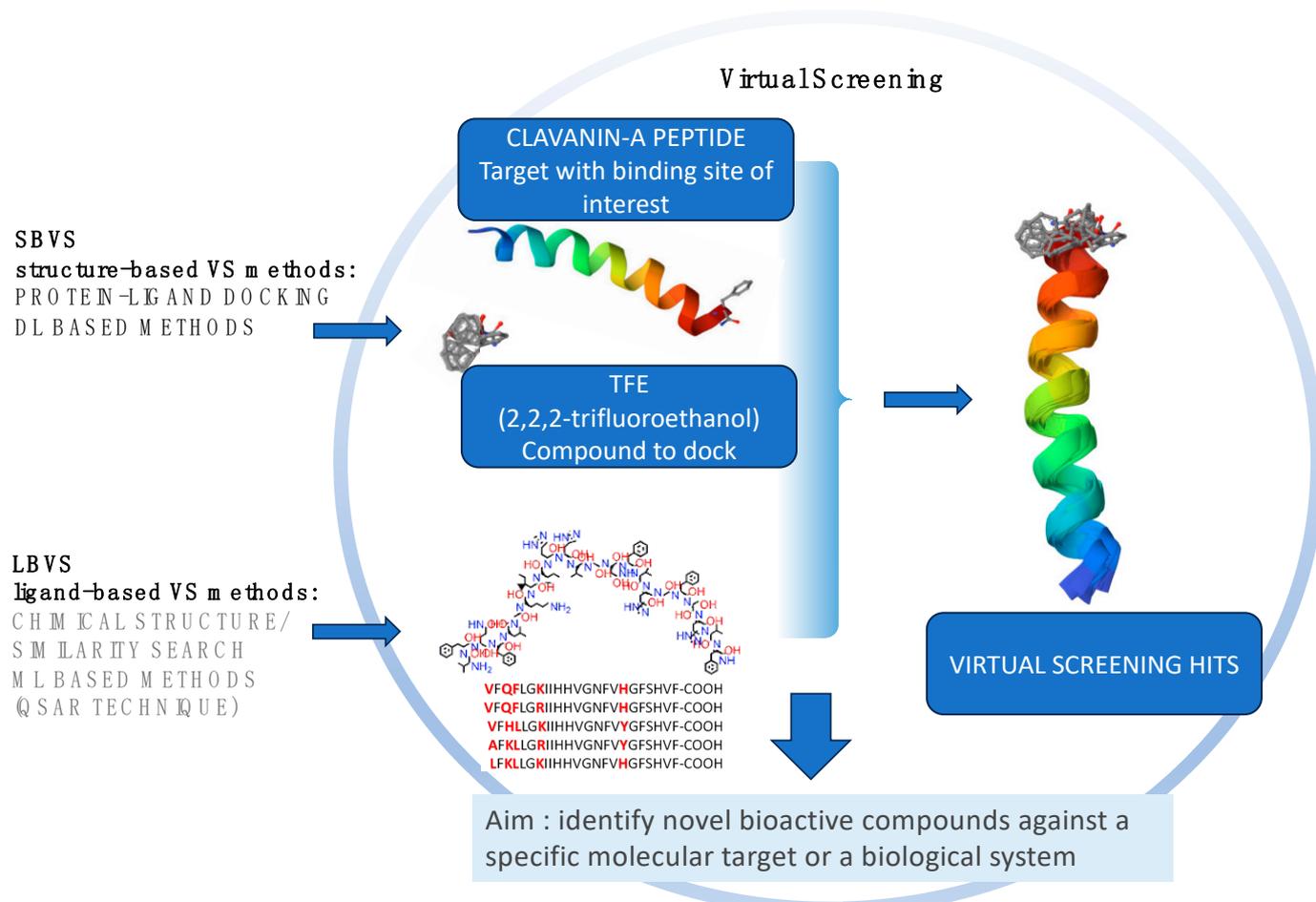
Drug toxicity prediction is one of the subtasks of drug discovery, and it allows us to predict how much a molecule could adversely affect humans. The use of AI technologies allows the testing of a specific molecule's toxicity, avoiding animal tests and high costs (Figure 2) [103,104]. The physico-chemical properties of molecules are another essential feature to assess in drug discovery studies. Their knowledge allows the understanding and modeling of the action of drugs. Among the numerous types of physico-chemical properties are the solubility, molar mass, charge, hydrophobicity, isoelectric point, and percentage of hydrophobic amino acids [105,106]. About 65% of the small molecule drugs are derived from natural products or their derivatives [107]; hence, drug bioactivity assessment has become an active area in drug discovery. AI techniques have been effectively applied to predicting drug bioactivities, such as anticancer, antiviral, and antibacterial activities (Figure 2) [108,109]. Among the different types of bioactive features, antimicrobial properties are gaining much attention as AMPs can be cutting-edge treatments for many infectious disorders. The effectiveness of AMPs against bacteria, fungi, and viruses has persisted for an extended period, making them the best option for addressing the growing problem of antibiotic resistance. Due to their wide-ranging actions, AMPs have become more prominent, particularly in therapeutic applications.



**Figure 2.** Schematic examples of the main tasks of pharmaceutical analysis and their possible resolution by AI techniques.

The prediction of AMPs has become difficult for academics due to the explosive increase of AMPs documented in databases. Wet-lab investigations to find antimicrobial peptides are exceedingly costly, time-consuming, and even impossible for some species. Therefore, to choose the optimal AMP candidates before *in vitro* trials, an efficient computational method must be developed [110]. Bioactive compounds can also assert an essential role as anticancer molecules. Indeed, most of them exert their antiproliferative effects by inhibiting different signaling pathways or intervening in cell-cycle arrest [111]. Various AI-based tools have recently been developed to help solve this aim. Another important subtask of drug discovery is the accurate binding prediction between a major histocompatibility complex (MHC) allele and bioactive peptides, as these last molecules are essential players in the synthesis of personalized cancer vaccines [112]. The immune system struggles to distinguish between a cancerous and a healthy cell. In a patient who has cancer with a particular MHC allele, only those peptides that bind with the MHC allele with high affinity help the immune system recognize the cancerous cells. AI approaches can help predict MHC-II binding, a fast alternative to wet-laboratory investigations, since experiments for MHC class II binding peptide identification are expensive and time consuming.

Finally, the protein structure is another important aspect to investigate in drug discovery as it allows us to understand the structural interactions, investigating the potential functionality of a protein and acquiring information on important binding domains for target proteins. Indeed, most drug targets are proteins that play essential roles in enzymatic activities, cell signaling, and cell–cell transduction. Although conventional experimental techniques, such as X-ray crystallography, cryogenic electron microscopy, and nuclear magnetic resonance spectroscopy have been proposed to investigate potential protein structures, they are still time-consuming and costly. Therefore, AI technology can help develop novel methods to fill the gap between the number of protein sequences and known protein structures. Virtual Screening (VS) is a computational approach that allows the prediction of the 3D structure of a chemical compound or a potential bioactive molecule against a specific target (Figure 3). It can be divided into ligand-based and structure-based methods. The former is used when very little structural information is available for the target and a set of active ligand molecules is known. Ligand-based methods include pharmacophore modeling and quantitative structure–activity relationship (QSAR) methods. The latter is used to model the interaction between a small molecule with a target protein at the atomic level, thus characterizing the behavior of small molecules in the binding site of target proteins and elucidating fundamental biochemical processes [113].

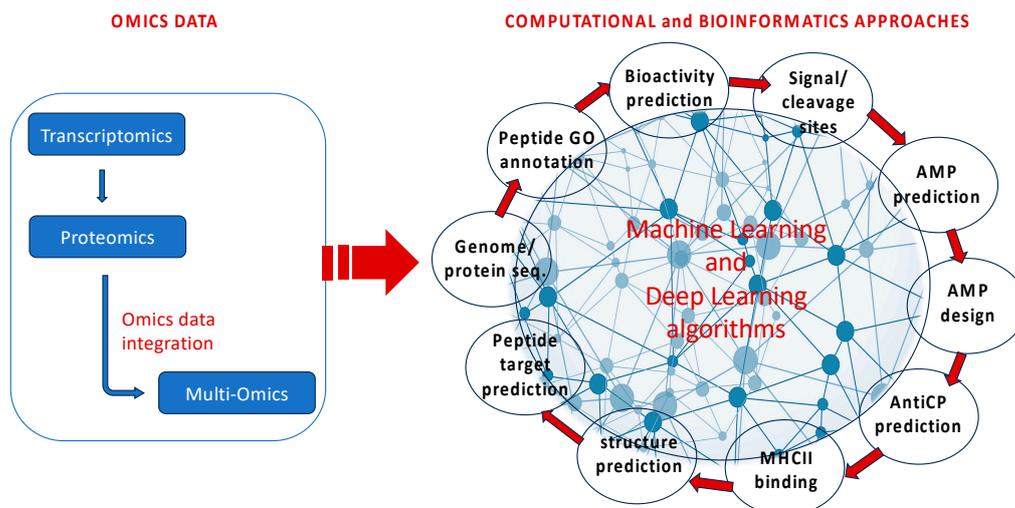


**Figure 3.** Structure-based and ligand-based VS approaches in protein docking.

### 5. AI-Based Web Tools for Bioactive Compound Identification and Analysis

As omics technologies have led to a vast increase in biological knowledge, this raises the necessity of figuring out how to mine helpful knowledge from it, which requires sophisticated data analysis and data-mining methods. To this aim, high-performance computing analysis methods have been developed to support the massive amount of data produced and, at the same time, give decision-making support to clinical and translational research [114,115].

By using ML algorithms, computers can learn without explicitly being programmed. It accomplishes prediction and classification tasks through pattern detection without using defined rules. Many bioactive compound analysis and identification methods use ML [116]. Several types of omics data, such as transcriptomics and proteomics, can be used alone or can be integrated into a multi-omics approach to be further analyzed by different computational methods based on AI (Figure 4). Different steps can be completed starting from genome or proteomic sequences to identify and analyze different features of potential bioactive molecules.



**Figure 4.** Several types of omics data can be used alone or in an integrated multi-omics approach to be further analyzed by different computational methods based on AI.

Indeed, different tools have been developed based on AI (Table 2):

#### 5.1. To predict the Antimicrobial Activity of Peptides

CAMP3 (<http://www.camp3.bicnirrh.res.in/>, accessed on 1 October 2023) (Collection of Antimicrobial Peptides) [117,118] allows the expansion and acceleration of AMP family-based studies. It uses information on the conserved sequence signatures captured as patterns and Hidden Markov Models (HMMs) [90]; AmPEPpy is a web tool for predicting AMP sequences using a random forest classifier (<https://github.com/tlawrence3/amPEPpy>, accessed on 1 October 2023) [119]; InverPrep contains the CALCAMPI algorithm that can calculate the physico-chemical properties (molar mass, charge, hydrophobicity, Boman index, aliphatic index, isoelectric point, and percentage of hydrophobic amino acids) of multiple peptides simultaneously ([http://ciencias.medellin.unal.edu.co/gruposdeinvestigacion/prospeccionydisenobiomoleculas/InverPep/public/home\\_en](http://ciencias.medellin.unal.edu.co/gruposdeinvestigacion/prospeccionydisenobiomoleculas/InverPep/public/home_en)) accessed on 2 October 2023 [120]; AntiBP2 (<http://crdd.osdd.net/raghava/antibp2/>, accessed on 2 October 2023) [121] is based on the support vector machine (SVM) algorithm using the composition of peptide sequences. The overall accuracy of the web server is about 92%, and the source of antibacterial peptides can also be predicted; AMPA is a web tool for the prediction of protein antimicrobial regions (<http://tcoffee.org.cat/apps/ampa>, accessed on 4 October 2023), and its main application is the fast automatic detection of antimicrobial regions in proteins that can serve as new templates for AMP design. AMPA-derived AI values can be used to classify proteins or domains as antimicrobial or non-antimicrobial automatically and compare different protein sequences in this regard. When used in conjunction with the T-coffee alignment tool, antimicrobial regions can be checked to identify potentially conserved antimicrobial domains [122].

#### 5.2. To Identify Peptides with Anticancer Properties

The AntiCP tool (<https://webs.iiitd.edu.in/raghava/anticp/>, accessed on 4 October 2023) [123] is a web-based prediction server based on machine learning techniques such as SVMs. It can predict every possible single-mutant/analog of a given peptide, and it can also predict their anticancer activity along with all the essential physico-chemical properties like hydrophobicity, charge, isoelectric point, etc. Other anticancer prediction tools using ML or DL models have been developed. Between them is a novel meta-approach which implements a user-friendly webserver for accurately identifying ACPs, which is called MLACP 2.0 [124]. The tool employs 11 different encoding schemes and eight different classifiers, including convolutional neural networks, to create a stable meta-

model (<https://balalab-skku.org/mlacp2>, accessed on 6 October 2023); ACP-MCAM can automatically learn adaptive embedding and the context sequence features of ACP [125]. Chen et al. used the features of amino acid dipeptide composition and pseudo-amino acid composition, combined with an SVM, to construct an ACP prediction algorithm called iACP ([http://crdd.osdd.net/raghava/anticp/multi\\_pep.php](http://crdd.osdd.net/raghava/anticp/multi_pep.php), accessed on 6 October 2023) [126].

Other methods can discriminate between ACPs and non-ACPs, and they include ACP [127], iACP-GAEnsC [128], TargetACP [129], ACPred [130], ACPred-FL [131], ACPred-Fuse [132], ACP-DL [133] and iACP-FSCM [134].

### 5.3. To Predict Peptide Binding with Immune Protein Classes

The automated platform to benchmark peptide–MHC class II binding prediction tools, called the Immune Epitope Database (IEDB), includes different prediction servers, such as NetMHCII 2.3 for T cell epitope prediction, along with other B cell epitope prediction tools (<http://tools.iedb.org/main/>, accessed on 10 October 2023) [135]. NetMHCII 2.3 (<https://services.healthtech.dtu.dk/services/NetMHCII-2.3/>, accessed on 10 October 2023) [136,137] uses artificial neuron networks (ANNs) to predict the binding of peptides to HLA-DR, HLA-DQ, HLA-DP, and mouse MHC class II alleles. The user is also guided in their choice of binding strength by strong and weak binding indicated in the output. Other computational methods for the prediction of MHC class II binding include ARB (<http://epitope.liai.org:8080/matrix>, accessed on 11 October 2023) [138], MHCpred (<http://SVRMHC.umn.edu/SVRMHCdb>, accessed on 11 October 2023) [139], TEPITOPE (<https://github.com/dmnfarrell/epitopepredict>, accessed on 11 October 2023) [140], and several others [141].

### 5.4. To Predict AMPs

Prediction tools can be applied to protein sequences from the proteome. A peptide must have many different features to be considered potentially bioactive, such as physico-chemical characteristics, signal peptides, and the location of their cleavage sites. Web tools such as SignalP 5.0 can achieve this aim (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>, accessed on 13 October 2023) [142]. This web tool can predict the presence of signal peptides or the location of their cleavage sites in proteins from different species through evolution, such as Archaea, Gram-positive Bacteria, Gram-negative Bacteria, and Eukarya. This method uses a combination of several ANNs and HMMs to predict cleavage sites and signal peptides/non-signal peptides. Both ANN and HMM are algorithms belonging to a subset of ML algorithms which are DL models. In ANN, the nodes of the networks are considered to be the neurons of the brain, and the edges connecting the nodes are considered to be “synapses”. DL models have also been applied among AMP predictors. Indeed, a key issue concerning DL models in AMP prediction is the need for samples in the positive class and their ambiguity in the negative class [106]. MultiPep (<https://agbg.shinyapps.io/MultiPep/>, accessed on 15 October 2023) is a hierarchical DL approach to the multi-label classification of peptide bioactivities [143]. PeptideRanker (<http://distilldeep.ucd.ie/PeptideRanker/>, accessed on 15 October 2023) [144] predicts peptide bioactivity using bioactivity probability scores. PeptideLocator (<http://bioware.ucd.ie/>, accessed on 15 October 2023) [145] is based on a BRNN algorithm. Other DL approaches used for different types of peptide analyses are Antimicrobial Peptide Scanner vr.2 (<https://www.dveltri.com/ascan/v2/ascan.html>, accessed on 16 October 2023) [146] and Deep-AmPEP30 (<https://cbbio.cis.um.edu.mo/AxPEP>, accessed on 16 October 2023) [147]. Antimicrobial Peptide Scanner vr.2 proposes a neural network model with convolutional and recurrent layers that leverage primary sequence composition to predict AMP properties.

A recent work by Hussain [148] describes an AMP prediction tool called AMP-PFPDeep, which is based on a deep neural network. It improves the accuracy of short antimicrobial peptide prediction using three different sequence encodings and an NN algorithm. The different sequences of the benchmark datasets used in the study were converted

into three-channel images comprising information related to the position, frequency, and sum of 12 physiochemical features as the first, second, and third channels, respectively.

AMPGANv2 is a bidirectional conditional generative adversarial network-based approach for rational AMP design (<https://gitlab.com/vail-uvm/amp-gan>, accessed on 18 October 2023). AMPGAN v2 uses generator–discriminator dynamics to learn data-driven priors and control generation using conditioning variables [149]. The bidirectional component, implemented using a learned encoder to map data samples into the latent space of the generator, aids in the iterative manipulation of candidate peptides.

### 5.5. To Predict Secondary Protein Structure

Web tools that use DL include the PSIPRED Protein Structure Prediction Server from the University College London Bioinformatics Unit (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>, accessed on 20 October 2023) [150], which consists of a two-stage neural network and is based on the position-specific scoring matrices generated by PSI-BLAST. It predicts highly accurate secondary structure predictions.

Another secondary structure prediction method is PSSpred (Protein Secondary Structure prediction) (<https://zhanggroup.org/PSSpred/>, accessed on 20 October 2023). It also uses PSI-BLAST to collect multiple sequence alignments. Then, amino-acid frequency and log-odds data are used to train the secondary structure, based on the Rumelhart error backpropagation method, and finally, a consensus of seven neural network predictors is used to predict the secondary structure prediction [151]. Among secondary structure prediction tools using DL approaches, there is also the APSSP2: Advanced Protein Secondary Structure Prediction Server [152] <http://crdd.osdd.net/raghava/apssp2/> (accessed on 20 October 2023).

### 5.6. To 3D Modeling

The AlphaFold method (<https://alphafold.ebi.ac.uk/>, accessed on 21 October 2023) [153] is a program that performs predictions of 3D protein structures and is designed as a DL system. IntFOLD (Integrated Fold Recognition) (<https://www.reading.ac.uk/bioinf/IntFOLD/>, accessed on 21 October 2023) is a fully automated, integrated pipeline for the prediction of 3D structures and functions from amino acid sequences [154]. RaptorX is another DL-based web server predicting both secondary predictions and 3D modeling ([raptorx.uchicago.edu](http://raptorx.uchicago.edu), accessed on 24 October 2023) [155]. ESyPred3D is an automated homology modeling program. Alignments are obtained by combining, weighing, and screening the results of several multiple alignment programs. The final three-dimensional structure is built using the MODELLER modeling package [156] (<http://www.fundp.ac.be/urbm/bioinfo/esypred/>, accessed on 24 October 2023).

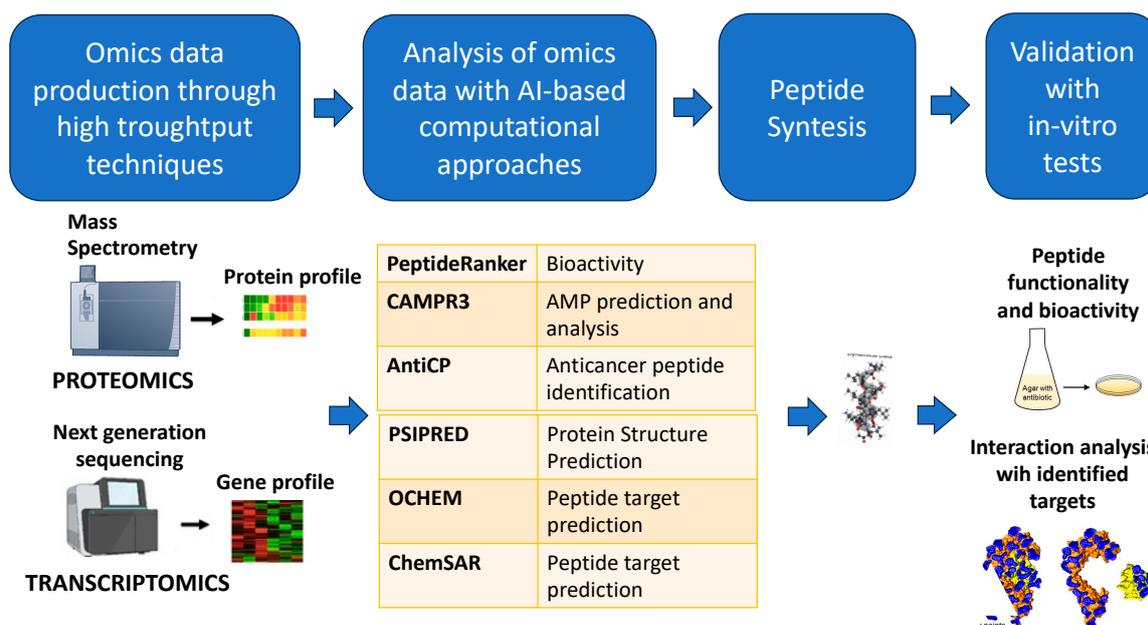
### 5.7. Web Tools Employed in VS Techniques

Docking and molecular modeling allow us to know information about molecule orientation and the spatial conformation of the molecule under investigation to infer potential interactions with specific targets as proteins involved in immunity [157]. Some examples of ML-based web servers are OCHEM (<https://ochem.eu/home/show.do>, accessed on 30 October 2023) [158–161] and ChemSAR (<http://chemsar.scbdd.com>, accessed on 30 October 2023) [162], which are both employed in target prediction.

Finally, we show one example of a pipeline used in the study of bioactive molecules with high-throughput omics technologies. Starting from transcriptomics or proteomics of a whole dataset, candidate peptide can be identified and further analyzed through different bioinformatics web tools that allow the evaluation the bioactivity of the putative peptide sequences, the AMP, and the anticancer activity. Moreover, the protein structure of putative peptides and the interaction with target proteins can be predicted. The candidate peptides identified can be then synthesized, and finally, *in vitro* testing can be used to validate the *in silico* predictions (Figure 5).

Table 2. AI-based tools developed for bioactive peptide analyses.

Drug Discovery Task	Web Tool/Method	AI Algorithm	Ref.	URL
Antimicrobial activity prediction	CAMP3	HMM (ML)	[118]	<a href="http://www.camp3.bicnirrh.res.in/">http://www.camp3.bicnirrh.res.in/</a>
	amPEPpy	RF (ML)	[119]	<a href="https://github.com/tlawrence3/amPEPpy">https://github.com/tlawrence3/amPEPpy</a>
	InverPrep		[120]	<a href="https://InverPep/public/home_en">https://InverPep/public/home_en</a>
	AntiBP2	SVM (ML)	[121]	<a href="https://webs.iiitd.edu.in/raghava/antibp2/index.html">https://webs.iiitd.edu.in/raghava/antibp2/index.html</a>
	AMPA	ML	[122]	<a href="http://tcoffee.crg.cat/apps/ampa">http://tcoffee.crg.cat/apps/ampa</a>
	Antimicrobial Peptide Scanner vr.2	ANN (DL)	[146]	<a href="https://www.dveltri.com/ascan/v2/ascan.html">https://www.dveltri.com/ascan/v2/ascan.html</a>
AMP design	Deep-AmPEP30	CNN (DL)	[147]	<a href="https://cbbio.cis.um.edu.mo/AxPEP">https://cbbio.cis.um.edu.mo/AxPEP</a>
	AMP-PFPDeep	DNN (DL)	[148]	<a href="https://github.com/WaqarHusain/sAMP-PFPDeep">https://github.com/WaqarHusain/sAMP-PFPDeep</a>
AMP design	AMPGANv2	generative adversarial network (DL)	[149]	<a href="https://gitlab.com/vail-uvvm/amp-gan">https://gitlab.com/vail-uvvm/amp-gan</a>
Anticancer properties	AntiCP	SVM (ML)	[123]	<a href="https://webs.iiitd.edu.in/raghava/anticp/">https://webs.iiitd.edu.in/raghava/anticp/</a>
	MLACP 2.0	ANN (DL)	[124]	<a href="https://balalab-skku.org/mlacp2">https://balalab-skku.org/mlacp2</a>
	ACP-MCAM	multi-kernel CNN	[125]	
	iACP	SVM (ML)	[126]	<a href="http://lin.uestc.edu.cn/server/iACP">http://lin.uestc.edu.cn/server/iACP</a>
	ACPP	SVM (ML)	[127]	<a href="https://github.com/brsaran/ACPP">https://github.com/brsaran/ACPP</a>
	iACP-GAEnsC	Genetic Algorithm (ML)	[128]	<a href="https://github.com/MLBC-lab/iACP-RF">https://github.com/MLBC-lab/iACP-RF</a>
	TargetACP	SVM (ML)	[129]	
	ACPred	RF+SVM (ML)	[130]	<a href="https://github.com/chaninlab/acpred-webserver">https://github.com/chaninlab/acpred-webserver</a>
	ACPred-FL	SVM (ML)	[131]	<a href="http://server.malab.cn/ACPred-FL">http://server.malab.cn/ACPred-FL</a>
	ACPred-Fuse	RF (ML)	[132]	<a href="http://server.malab.cn/ACPred-Fuse">http://server.malab.cn/ACPred-Fuse</a>
ACP-DL and iACP-FSCM	LSTM (DL) Genetic Algorithm	[133] [134]	<a href="https://github.com/haichengyi/ACP-DL">https://github.com/haichengyi/ACP-DL</a> <a href="http://camt.pythonanywhere.com/iACP-FSCM">http://camt.pythonanywhere.com/iACP-FSCM</a>	
MHC class II binding prediction	NetMHCII 2.3	ANN (DL)	[135]	<a href="https://services.healthtech.dtu.dk/services/NetMHCII-2.3/">https://services.healthtech.dtu.dk/services/NetMHCII-2.3/</a>
	ARB	Average Relative Binding (ARB) matrix PLS-based	[138]	<a href="http://epitope.liai.org:8080/matrix">http://epitope.liai.org:8080/matrix</a>
	MHCpred	allele-specific multivariate statistical model	[139]	<a href="http://SVRMHC.umn.edu/SVRMHCdb">http://SVRMHC.umn.edu/SVRMHCdb</a>
	TEPITOPEpan	ANN (DL)	[140]	<a href="http://www.biokdd.fudan.edu.cn/Service/TEPITOPEpan/">http://www.biokdd.fudan.edu.cn/Service/TEPITOPEpan/</a>
Signal peptides cleavage site location	SignalP 5.0	ANN + HMM (ML + DL)	[142]	<a href="https://services.healthtech.dtu.dk/services/SignalP-5.0/">https://services.healthtech.dtu.dk/services/SignalP-5.0/</a>
Peptide bioactivity	MultiPep	deep neural network multi-label	[143]	<a href="https://agbg.shinyapps.io/MultiPep/">https://agbg.shinyapps.io/MultiPep/</a>
	PeptideRanker	feed-forward neural network (DL)	[144]	<a href="http://distilldeep.ucd.ie/PeptideRanker/">http://distilldeep.ucd.ie/PeptideRanker/</a>
	PeptideLocator	BRNN (DL)	[145]	<a href="http://bioware.ucd.ie/">http://bioware.ucd.ie/</a>
Secondary structure prediction	PSIPRED	Two-stage NN (DL)	[150]	<a href="http://bioinf.cs.ucl.ac.uk/psipred/psiform.html">http://bioinf.cs.ucl.ac.uk/psipred/psiform.html</a>
	PSSpred	Neural Network (DL)	[151]	<a href="https://zhanggroup.org/PSSpred/">https://zhanggroup.org/PSSpred/</a>
	ASPPS2	Neural Network (DL)	[152]	<a href="http://crdd.osdd.net/raghava/apssp2/">http://crdd.osdd.net/raghava/apssp2/</a>
3D modeling	AlphaFold	HMM (DL)	[153]	<a href="https://alphafold.ebi.ac.uk/">https://alphafold.ebi.ac.uk/</a>
	IntFOLD	Neural Network (DL)	[154]	<a href="https://www.reading.ac.uk/bioinf/IntFOLD/">https://www.reading.ac.uk/bioinf/IntFOLD/</a>
	RaptorX	Neural Network (DL)	[155]	<a href="https://raptorx.uchicago.edu/">https://raptorx.uchicago.edu/</a>
	ESyPred3D	Neural Network (DL)	[156]	<a href="http://www.fundp.ac.be/urbm/bioinfo/esypred/">http://www.fundp.ac.be/urbm/bioinfo/esypred/</a>
Virtual Screening	OCHEM	Neural Network	[161]	<a href="https://ochem.eu/home/show.do">https://ochem.eu/home/show.do</a>
	ChemSAR	ML	[162]	<a href="http://chemsar.scbdd.com">http://chemsar.scbdd.com</a>



**Figure 5.** Example of a pipeline used in the study of bioactive molecules with high-throughput omics technologies.

## 6. Conclusions

Drug discovery has always been a complex and time-consuming endeavor that traditionally relies on labor-intensive techniques, such as trial-and-error experimentation and high-throughput screening. However, these methods can be slow, costly, and often yield results with low accuracy. AI techniques such as ML and natural language processing, combined with the experimental production of omics data, offer the potential to accelerate and improve this process by enabling more efficient and accurate analyses of large amounts of data. This combined approach allows researchers to examine a large number of potential drug compounds to identify those with the desired properties. Classic methods can be limited by the availability of suitable test compounds and the difficulty of accurately predicting their pharmacological behavior. Different algorithms based on AI, including supervised and unsupervised learning methods, reinforcement, and evolutionary or rule-based algorithms, can potentially contribute to solving these problems. For instance, the efficacy and toxicity of new drug compounds can be predicted using these approaches with greater accuracy and efficiency than when using traditional methods. Furthermore, AI-based algorithms can also be employed to identify new targets for drug development, such as the specific proteins or genetic pathways involved in diseases. This can expand the scope of drug discovery beyond the limitations of more conventional approaches and may lead to the development of novel and more effective drugs. AI-based methods, on the other hand, can improve the efficiency and accuracy of drug discovery processes and lead to the development of more effective drugs. Furthermore, high-throughput gene sequencing has revolutionized the method used to identify novel molecular targets for drug discovery.

Since there has been increasing attention to finding new solutions for a sustainable supply of bioactive compounds, we would like evidence of the most recent bioinformatics methods to be connected to bioactive compound research. We also propose that a promising source of bioactive molecules, such as ACPs, antiviral, antimicrobial, anti-oxidative stress and immunomodulatory compounds, along with innovative solutions as therapeutic strategies, might come from the study of ascidian innate immunity processes.

Finally, the research's expected impacts are therefore multiple: from the standardization of a workflow of several techniques that can be replicated by using omics data coming from the study of immune processes of different animal models and different AI based algorithms to the identification of various classes of bioactive molecules that can form the

basis for a new groups of drugs which are pharmacologically more efficient and have fewer side effects.

**Author Contributions:** Conceptualization, L.L.P. and A.V.; methodology, L.L.P. and A.V.; data curation, L.L.P., A.V., M.V., A.U., V.A. and M.M.; writing—original draft preparation, L.L.P. and A.V.; writing—review and editing, L.L.P. and A.V.; supervision, L.L.P. and A.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zeng, L.; Swalla, B.J. Molecular phylogeny of the protochordates: Chordate evolution. *Can. J. Zool.* **2005**, *83*, 24–33. [[CrossRef](#)]
2. Delsuc, F.; Brinkmann, H.; Chourrout, D.; Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **2006**, *439*, 965–968. [[CrossRef](#)] [[PubMed](#)]
3. Tsagkogeorga, G.; Turon, X.; Hopcroft, R.R.; Tilak, M.K.; Feldstein, T.; Shenkar, N.; Loya, Y.; Huchon, D.; Douzery, E.J.P.; Del-suc, F. An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol. Biol.* **2009**, *9*, 187. [[CrossRef](#)] [[PubMed](#)]
4. Delsuc, F.; Philippe, H.; Tsagkogeorga, G.; Simion, P.; Tilak, M.-K.; Turon, X.; López-Legentil, S.; Piette, J.; Lemaire, P.; Douzery, E.J.P. A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* **2018**, *16*, 39. [[CrossRef](#)] [[PubMed](#)]
5. Parrinello, N. The reaction of *Ciona intestinalis* L. to subcuticular erythrocyte and protein injection. *Dev. Comp. Immunol.* **1981**, *5*, 105–110. [[CrossRef](#)]
6. Parrinello, N.; Patricolo, E.; Canicatti, C. Inflammatory-like reaction in the tunic of *Ciona intestinalis* (Tunicata). Encapsulation and tissue injury I. *Biol. Bull.* **1984**, *167*, 229–237. [[CrossRef](#)]
7. Parrinello, N.; Patricolo, E.; Canicatti, C. Inflammatory-like reaction in the tunic of *Ciona intestinalis* (Tunicata). Encapsulation tissue injury II. *Biol. Bull.* **1984**, *167*, 238–250. [[CrossRef](#)]
8. Vizzini, A.; Pergolizzi, M.; Vazzana, M.; Salerno, G.; Di Sano, C.; Macaluso, P.; Arizza, V.; Parrinello, D.; Cammarata, M.; Parrinello, N. FACIT collagen (1alpha-chain) is expressed by hemocytes and epidermis during the inflammatory response of the ascidian *Ciona intestinalis*. *Dev. Comp. Immunol.* **2008**, *32*, 682–692. [[CrossRef](#)]
9. Zhao, Y.; Li, J. Excellent Chemical and Material Cellulose from Tunicates: Diversity in Cellulose Production Yield and Chemical and Morphological Structures from Different Tunicate Species. *Cellulose* **2014**, *21*, 3427–3441. [[CrossRef](#)]
10. Vizzini, A.; Parrinello, D.; Sanfratello, M.A.; Trapani, M.R.; Mangano, V.; Parrinello, N.; Cammarata, M. Upregulated transcription of phenoloxidase genes in the pharynx and endostyle of *Ciona intestinalis* in response to LPS. *J. Invertebr. Pathol.* **2015**, *126C*, 6–11. [[CrossRef](#)]
11. Zeng, F.; Wunderer, J.; Salvenmoser, W.; Ederth, T.; Rothbacher, U. Identifying Adhesive Components in a Model Tunicate. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20190197. [[CrossRef](#)] [[PubMed](#)]
12. Li, S.; Huang, X.; Chen, Y.; Li, X.; Zhan, A. Identification and Characterization of Proteins Involved in Stolon Adhesion in the Highly Invasive Fouling Ascidian *Ciona Robusta*. *Biochem. Biophys. Res. Commun.* **2019**, *510*, 91–96. [[CrossRef](#)]
13. Taylor, S.W.; Kammerer, B.; Bayer, E. New Perspectives in the Chemistry and Biochemistry of the Tunichromes and Related Compounds. *Chem. Rev.* **1997**, *97*, 333–346. [[CrossRef](#)] [[PubMed](#)]
14. Dou, X.; Dong, B. Origins and Bioactivities of Natural Compounds Derived from Marine Ascidiaceans and Their Symbionts. *Mar. Drugs* **2019**, *17*, 670. [[CrossRef](#)] [[PubMed](#)]
15. Casertano, M.; Menna, M.; Imperatore, C. The Ascidian-Derived Metabolites with Antimicrobial Properties. *Antibiotics* **2020**, *9*, 510. [[CrossRef](#)]
16. Adrian, T.E. Novel Marine-Derived Anti-Cancer Agents. *Curr. Pharm. Des.* **2007**, *13*, 3417–3426. [[CrossRef](#)]
17. Ramesh, C.; Tulasi, B.R.; Raju, M.; Thakur, N.; Dufossé, L. Marine Natural Products from Tunicates and Their Associated Microbes. *Mar. Drugs* **2021**, *19*, 308. [[CrossRef](#)]
18. Sivonen, K.; Leikoski, N.; Fewer, D.P.; Jokela, J. Cyanobactins—Ribosomal Cyclic Peptides Produced by Cyanobacteria. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 1213–1225. [[CrossRef](#)]
19. Zeng, F.; Wunderer, J.; Salvenmoser, W.; Hess, M.W.; Ladurner, P.; Rothbacher, U. Papillae Revisited and the Nature of the Adhesive Secreting Collocytes. *Dev. Biol.* **2019**, *448*, 183–198290. [[CrossRef](#)]
20. Mauro, M.; Lazzara, V.; Punginelli, D.; Arizza, V.; Vazzana, M. Antitumoral compounds from vertebrate sister group: A review of Mediterranean ascidiaceans. *Dev. Comp. Immunol.* **2020**, *108*, 103669. [[CrossRef](#)]
21. Akira, S.; Uematsu, S.; Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **2006**, *124*, 783–801. [[CrossRef](#)] [[PubMed](#)]
22. Beutler, B.; Jiang, Z.; Georgel, P.; Crozat, K.; Croker, B.; Rutschmann, S.; Du, X.; Hoebe, K. Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* **2006**, *24*, 353–389. [[CrossRef](#)] [[PubMed](#)]
23. Nathan, C. Points of control in inflammation. *Nature* **2002**, *420*, 846–852. [[CrossRef](#)] [[PubMed](#)]
24. Medzhitov, R. Origin and physiological roles of inflammation. *Nature* **2008**, *454*, 428–435. [[CrossRef](#)] [[PubMed](#)]

25. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **2007**, *449*, 819–826. [[CrossRef](#)] [[PubMed](#)]
26. Ishii, K.J.; Koyama, S.; Nakagawa, A.; Coban, C.; Akira, S. Host innate immune receptors and beyond: Making sense of microbial infections. *Cell Host Microbe* **2008**, *3*, 352–363. [[CrossRef](#)]
27. Beutler, B.A. TLRs and innate immunity. *Blood* **2009**, *113*, 1399–1407. [[CrossRef](#)]
28. Vizzini, A.; Parisi, M.G.; Di Falco, F.; Cardinale, L.; Cammarata, M.; Arizza, V. Identification of CPE and GAIT elements in 3'UTR of macrophage migration inhibitory factor (MIF) involved in the inflammatory response induced by LPS in *Ciona robusta*. *Mol. Immunol.* **2018**, *99*, 66–74. [[CrossRef](#)]
29. Vizzini, A. Gene expression and regulation of molecules involved in pharynx inflammatory response induced by LPS in *Ciona intestinalis*. *Invertebr. Surviv. J.* **2017**, *14*, 119–128.
30. Arizza, V.; Bonura, A.; La Paglia, L.; Urso, A.; Pinsino, A.; Vizzini, A. Transcriptional and in silico analyses of MIF cytokine and TLR signalling interplay in the LPS inflammatory response of *Ciona robusta*. *Sci. Rep.* **2020**, *10*, 11339. [[CrossRef](#)]
31. Vizzini, A.; Bonura, A.; La Paglia, L.; Fiannaca, A.; La Rosa, M.; Urso, A.; Arizza, V. ceRNA Network Regulation of TGF- $\beta$ , WNT, FOXO, Hedgehog Pathways in the Pharynx of *Ciona robusta*. *Int. J. Mol. Sci.* **2021**, *22*, 3497. [[CrossRef](#)] [[PubMed](#)]
32. Vizzini, A.; Bonura, A.; La Paglia, L.; Fiannaca, A.; La Rosa, M.; Urso, A.; Mau-ro, M.; Vazzana, M.; Arizza, V. Transcriptomic Analyses Reveal 2 and 4 Family Members of Cytochromes P450 (CYP) Involved in LPS Inflammatory Response in Pharynx of *Ciona robusta*. *Int. J. Mol. Sci.* **2021**, *22*, 11141. [[CrossRef](#)] [[PubMed](#)]
33. Vizzini, A.; Bonura, A.; Parrinello, D.; Sanfratello, M.A.; Longo, V.; Colombo, P. LPS Challenge Regulates Gene Expression and Tissue Localization of a *Ciona intestinalis* Gene through an Alternative Polyadenylation Mechanism. *PLoS ONE* **2013**, *8*, e63235. [[CrossRef](#)] [[PubMed](#)]
34. Ohtsuka, Y.; Inagaki, H. In silico identification and functional validation of linear cationic  $\alpha$ -helical antimicrobial peptides in the ascidian *Ciona intestinalis*. *Sci. Rep.* **2020**, *10*, 12619. [[CrossRef](#)]
35. Giacomelli, S.; Melillo, D.; Lambris, J.D.; Pinto, M.R. Immune competence of the *Ciona intestinalis* pharynx: Complement system-mediated activity. *Fish. Shellfish. Immunol.* **2012**, *33*, 946–952. [[CrossRef](#)]
36. Parrinello, N.; Vizzini, A.; Arizza, V.; Salerno, G.; Parrinello, D.; Cammarata, M.; Giaramita, F.T.; Vazzana, M. Enhanced expression of a cloned and sequenced *Ciona intestinalis* TNF alpha like (CiTNF alpha) gene during the LPS-induced inflammatory response. *Cell Tissue Res.* **2008**, *334*, 305–317. [[CrossRef](#)]
37. Vizzini, A.; Di Falco, F.; Parrinello, D.; Sanfratello, M.A.; Cammarata, M. Transforming growth factor beta (CiTGF-beta) gene expression is induced in the inflammatory reaction of *Ciona intestinalis*. *Dev. Comp. Immunol.* **2016**, *55*, 102–110. [[CrossRef](#)]
38. Vizzini, A.; Falco, F.D.; Parrinello, D.; Sanfratello, M.A.; Mazarella, C.; Parrinello, N.; Cammarata, M. *Ciona intestinalis* interleukin 17-like genes expression is upregulated by LPS challenge. *Dev. Comp. Immunol.* **2015**, *48*, 129–137. [[CrossRef](#)]
39. Brodin, P.; Davis, M.M. Human immune system variation. *Nat. Rev. Immunol.* **2017**, *17*, 21–29. [[CrossRef](#)]
40. Parkin, J.; Cohen, B. An overview of the immune system. *Lancet* **2001**, *357*, 1777–1789. [[CrossRef](#)]
41. Chaplin, D.D. Overview of the immune response. *J. Allergy Clin. Immunol.* **2010**, *125*, S3–S23. [[CrossRef](#)] [[PubMed](#)]
42. Iwasaki, A.; Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nat. Immunol.* **2004**, *5*, 987–995. [[CrossRef](#)] [[PubMed](#)]
43. Chakrabarti, S.; Guha, S.; Majumder, K. Food-Derived Bioactive Peptides in Human Health: Challenges and Opportunities. *Nutrients* **2018**, *10*, 1738. [[CrossRef](#)] [[PubMed](#)]
44. Segura-Campos, M.; Chel-Guerrero, L.; Betancur-Ancona, D.; Hernandez-Escalante, V.M. Bioavailability of Bioactive Peptides. *Food Rev. Int.* **2011**, *27*, 213–226. [[CrossRef](#)]
45. Manikkam, V.; Vasiljevic, T.; Donkor, O.N.; Mathai, M.L. A Review of Potential Marine-derived Hypotensive and Anti-obesity Peptides. *Crit. Rev. Food Sci. Nutr.* **2016**, *56*, 92–112. [[CrossRef](#)] [[PubMed](#)]
46. Kitts, D.D.; Weiler, K. Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery. *Curr. Pharm. Des.* **2003**, *9*, 1309–1323. [[CrossRef](#)] [[PubMed](#)]
47. Fiat, A.M.; Migliore-Samour, D.; Jolles, P.; Drouet, L.; Bal dit Sollier, C.; Caen, J. Biologically active peptides from milk proteins with emphasis on two examples concerning antithrombotic and immunomodulating activities. *J. Dairy Sci.* **1993**, *76*, 301–310. [[CrossRef](#)] [[PubMed](#)]
48. Fagundes, P.C.; Ceotto, H.; Potter, A.; Vasconcelos de Paiva Brito, M.A.; Brede, D.; Nes, I.F.; Bastos Mdo, C. Hyicin; a bioactive peptide produced by *Staphylococcus hyicus* with potential applications for food preservation. *Res. Microbiol.* **2011**, *162*, 1052–1059. [[CrossRef](#)]
49. Ayuningrum, D.; Liu, Y.; Riyanti; Sibero, M.T.; Kristiana, R.; Asagabaldan, M.A.; Wuisan, Z.G.; Trianto, A.; Radjasa, O.K.; Sabdon, A.; et al. Tunicate-Associated Bacteria Show a Great Potential for the Discovery of Antimicrobial Compounds. *PLoS ONE* **2019**, *14*, e0213797. [[CrossRef](#)]
50. Chen, L.; Hu, J.-S.; Xu, J.-L.; Shao, C.-L.; Wang, G.-Y. Biological and Chemical Diversity of Ascidian-Associated Microorganisms. *Mar. Drugs* **2018**, *16*, 362. [[CrossRef](#)]
51. Azumi, K.; Yokosawa, H.; Ishii, S. Halocyanines: Novel Antimicrobial Tetrapeptide-like Substances Isolated from the Hemocytes of the Solitary Ascidian *Halocynthia roretzi*. *Biochemistry* **1990**, *29*, 159–165. [[CrossRef](#)] [[PubMed](#)]
52. Fedders, H.; Michalek, M.; Grötzinger, J.; Leippe, M. An Exceptional Salt-Tolerant Antimicrobial Peptide Derived from a Novel Gene Family of Haemocytes of the Marine Invertebrate *Ciona intestinalis*. *Biochem. J.* **2008**, *416*, 65–75. [[CrossRef](#)] [[PubMed](#)]

53. Fedders, H.; Leippe, M. A Reverse Search for Antimicrobial Peptides in *Ciona intestinalis*: Identification of a Gene Family Expressed in Hemocytes and Evaluation of Activity. *Dev. Comp. Immunol.* **2008**, *32*, 286–298. [[CrossRef](#)] [[PubMed](#)]
54. Jang, W.S.; Kim, K.N.; Lee, Y.S.; Nam, M.H.; Lee, I.H. Halocidin: A New Anti-microbial Peptide from Hemocytes of the Solitary Tunicate, *Halocynthia aurantium*. *FEBS Lett.* **2002**, *521*, 81–86. [[CrossRef](#)] [[PubMed](#)]
55. Menzel, L.P.; Lee, I.H.; Sjostrand, B.; Lehrer, R.I. Immunolocalization of Clavanins in *Styela clava* Hemocytes. *Dev. Comp. Immunol.* **2002**, *26*, 505–515. [[CrossRef](#)] [[PubMed](#)]
56. Lee, I.H.; Zhao, C.; Nguyen, T.; Menzel, L.; Waring, A.J.; Sherman, M.A.; Lehrer, R.I. Clavaspirin, an Antibacterial and Haemolytic Peptide from *Styela clava*. *J. Pept. Res.* **2001**, *58*, 445–456. [[PubMed](#)]
57. Lee, I.H.; Cho, Y.; Lehrer, R.I. Styelins, Broad-Spectrum Antimicrobial Peptides from the Solitary Tunicate, *Styela clava*. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **1997**, *118*, 515–521. [[CrossRef](#)] [[PubMed](#)]
58. Taylor, S.W.; Craig, A.G.; Fischer, W.H.; Park, M.; Lehrer, R.I.; Styelin, D. An Extensively Modified Antimicrobial Peptide from Ascidian Hemocytes. *J. Biol. Chem.* **2000**, *275*, 38417–38426. [[CrossRef](#)]
59. Tincu, J.A.; Menzel, L.P.; Azimov, R.; Sands, J.; Hong, T.; Waring, A.J.; Taylor, S.W.; Lehrer, R.I. Plicatamide, an Antimicrobial Octapeptide from *Styela plicata* Hemocytes. *J. Biol. Chem.* **2003**, *278*, 13546–13553. [[CrossRef](#)]
60. Galinier, R.; Roger, E.; Sautiere, P.E.; Aumelas, A.; Banaigs, B.; Mitta, G. Halocytin and papillosin, two new antimicrobial peptides isolated from hemocytes of the solitary tunicate, *Halocynthia papillosa*. *J. Pept. Sci.* **2009**, *15*, 48–55. [[CrossRef](#)]
61. Lu, Y.; Zhuang, Y.; Liu, J. Mining antimicrobial peptides from small open reading frames in *Ciona intestinalis*. *J. Pept. Sci.* **2014**, *20*, 25–29. [[CrossRef](#)] [[PubMed](#)]
62. Bonura, A.; Vizzini, A.; Vlah, S.; Gervasi, F.; Longo, A.; Melis, M.R.; Schildberg, F.A.; Colombo, P. Ci8 short, a novel LPS-induced peptide from the ascidian *Ciona intestinalis*, modulates responses of the human immune system. *Immunobiology* **2018**, *223*, 210–219. [[CrossRef](#)] [[PubMed](#)]
63. Longo, V.; Longo, A.; Martorana, A.; Lauria, A.; Augello, G.; Azzolina, A.; Cervello, M.; Colombo, P. Identification of an LPS-Induced Chemo-Attractive Peptide from *Ciona robusta*. *Mar. Drugs* **2020**, *18*, 209. [[CrossRef](#)] [[PubMed](#)]
64. Ballarin, L.; Cima, F.; Floreani, M.; Sabbadin, A. Oxidative Stress Induces Cytotoxicity during Rejection Reaction in the Compound Ascidian *Botryllus schlosseri*. *Comp. Biochem. Physiol. Part. C Toxicol. Pharmacol.* **2002**, *133*, 411–418. [[CrossRef](#)] [[PubMed](#)]
65. Gruber, C.W.; Muttenthaler, M. Discovery of defense- and neuropeptides in social ants by genome-mining. *PLoS ONE* **2012**, *7*, e32559. [[CrossRef](#)] [[PubMed](#)]
66. Belarmino, L.C.; Benko-Iseppon, A.M. Databank based mining on the track of antimicrobial weapons in plant genomes. *Curr. Protein Pept. Sci.* **2010**, *11*, 195–198. [[CrossRef](#)]
67. Silverstein, K.A.; Graham, M.A.; Paape, T.D.; VandenBosch, K.A. Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol.* **2005**, *138*, 600–610. [[CrossRef](#)]
68. Giacomelli, L.; Nanni, V.; Lenzi, L.; Zhuang, J.; Dalla Serra, M.; Banfield, M.J.; Town, C.D.; Silverstein, K.A.; Baraldi, E.; Moser, C. Identification and characterization of the defensin-like gene family of grapevine. *Mol. Plant Microbe Interact.* **2012**, *25*, 1118–1131.
69. Fedders, H.; Podschun, R.; Leippe, M. The antimicrobial peptide Ci-MAM-A24 is highly active against multidrug-resistant and anaerobic bacteria pathogenic for humans. *Int. J. Antimicrob. Agents* **2010**, *36*, 264. [[CrossRef](#)]
70. Arron, J.R.; Choi, Y. Osteoimmunology: Bone versus immune system. *Nature* **2000**, *408*, 535–536. [[CrossRef](#)]
71. Kim, J.; Bentley, P.J.; Aickelin, U.; Green-smith, J.; Tedesco, G.; Twycross, J. Immune system approaches to intrusion detection—a review. *Nat. Comput.* **2007**, *6*, 413–466. [[CrossRef](#)]
72. Marcus, A.; Gowen, B.G.; Thompson, T.W.; Iannello, A.; Ardolino, M.; Deng, W.; Wang, L.; Shifrin, N.; Raulet, D.H. Recognition of tumors by the innate immune system and natural killer cells. *Adv. Immunol.* **2014**, *122*, 91–128. [[CrossRef](#)] [[PubMed](#)]
73. Schepetkin, I.A.; Kirpotina, L.N.; Jakiw, L.; Khlebnikov, A.I.; Blaskovich, C.L.; Jutila, M.A.; Quinn, M.T. Immunomodulatory activity of oenothien B isolated from *Epilobium angustifolium*. *J. Immunol.* **2009**, *183*, 6754–6766. [[CrossRef](#)] [[PubMed](#)]
74. Kim, R.; Emi, M.; Tanabe, K. Cancer immunoediting from immune surveillance to immune escape. *Immunology* **2007**, *121*, 1–14. [[CrossRef](#)] [[PubMed](#)]
75. Hendry, S.A.; Farnsworth, R.H.; Solomon, B.; Achen, M.G.; Stacker, S.A.; Fox, S.B. The Role of the Tumor Vasculature in the Host Immune Response: Implications for Therapeutic Strategies Targeting the Tumor Microenvironment. *Front. Immunol.* **2016**, *7*, 621. [[CrossRef](#)] [[PubMed](#)]
76. Ozawa, A.; Lindberg, I.; Roth, B.; Kroeze, W.K. Deorphanization of Novel Peptides and Their Receptors. *Aaps J.* **2010**, *12*, 378–384. [[CrossRef](#)] [[PubMed](#)]
77. Kawada, T.; Osugi, T.; Matsubara, S.; Sakai, T.; Shiraishi, A.; Yamamoto, T.; Satake, H. Omics Studies for the Identification of Ascidian Peptides, Cognate Receptors, and Their Relevant Roles in Ovarian Follicular Development. *Front. Endocrinol. Lausanne* **2022**, *13*, 858885. [[CrossRef](#)]
78. Franchi, N.; Ballarin, L.; Cima, F. Botryllin, a Novel Antimicrobial Peptide from the Colonial Ascidian *Botryllus schlosseri*. *Mar. Drugs* **2023**, *21*, 74. [[CrossRef](#)]
79. Zhang, Z.; Wu, S.; Stenoien, D.L.; Paša-Tolić, L. High-throughput proteomics. *Annu. Rev. Anal. Chem. Palo Alto Calif.* **2014**, *7*, 427–454. [[CrossRef](#)]
80. Schmidt, A.; Forne, I.; Imhof, A. Bioinformatic analysis of proteomics data. *BMC Syst. Biol.* **2014**, *8* (Suppl. 2), S3. [[CrossRef](#)]

81. Matos, A.; Domínguez-Pérez, D.; Almeida, D.; Agüero-Chapin, G.; Campos, A.; Osório, H.; Vasconcelos, V.; Antunes, A. Shotgun Proteomics of Ascidiacs Tunic Gives New Insights on Host–Microbe Interactions by Revealing Diverse Antimicrobial Peptides. *Mar. Drugs* **2020**, *18*, 362. [[CrossRef](#)] [[PubMed](#)]
82. Litman, G.W.; Cannon, J.P.; Dishaw, L.J. Reconstructing immune phylogeny: New perspectives. *Nat. Rev. Immunol.* **2005**, *5*, 866–879. [[CrossRef](#)] [[PubMed](#)]
83. Wiebe, M.S.; Traktman, P. Poxviral B1 Kinase Overcomes Barrier to Autointegration Factor, a Host Defense against Virus Replication. *Cell Host Microbe* **2007**, *1*, 187–197. [[CrossRef](#)] [[PubMed](#)]
84. Liberti, A.; Pollastro, C.; Pinto, G.; Illiano, A.; Marino, R.; Amoresano, A.; Spagnuolo, A.; Sordino, P. Transcriptional and proteomic analysis of the innate immune response to microbial stimuli in a model invertebrate chordate. *Front. Immunol. Sec. Comp. Immunol.* **2023**, *14*, 1217077. [[CrossRef](#)] [[PubMed](#)]
85. Mukhamediev, R.I.; Popova, Y.; Kuchin, Y.; Zaitseva, E.; Kalimoldayev, A.; Symagulov, A.; Levashenko, V.; Abdoldina, F.; Gopejenko, V.; Yakunin, K.; et al. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics* **2022**, *10*, 2552. [[CrossRef](#)]
86. Rokach, L.; Maimon, O. Decision Trees. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2005. [[CrossRef](#)]
87. Van Messem, A. Chapter 10—Support vector machines: A robust prediction method with applications in bioinformatics. In *Handbook of Statistics*; Arni, S.R., Rao, S., Rao, C.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2020. [[CrossRef](#)]
88. Liao, J.G.; Chin, K.-V. Logistic regression for disease classification using microarray data: Model selection in a large  $p$  and small  $n$  case. *Bioinformatics* **2007**, *23*, 1945–1951. [[CrossRef](#)]
89. Zheng, J.; Rakovski, C. On the Application of Principal Component Analysis to Classification Problems. *Data Sci. J.* **2021**, *20*, 26. [[CrossRef](#)]
90. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. [[CrossRef](#)]
91. Nascimento, M.; Fonseca e Silva, F.; Sáfadi, T.; Campana Nascimento, A.C.; Ferreira, T.E.M.; Azevedo Barroso, L.M.; Feirreira Azevedo, C.; Faccione Guimarães, S.E.; Lopes Serão, N.V. Independent Component Analysis (ICA) based-clustering of temporal RNA-seq data. *PLoS ONE* **2017**, *12*, e0181195. [[CrossRef](#)]
92. Taebi, A. Deep Learning for Computational Hemodynamics: A Brief Review of Recent Advances. *Fluids* **2022**, *7*, 197. [[CrossRef](#)]
93. Bala, R.; Kumar, D. Classification Using ANN: A Review. *Int. J. Comput. Intell. Res.* **2017**, *13*, 1811–1820.
94. Aoki, G.; Sakakibara, Y. Convolutional neural networks for classification of alignments of non-coding RNA sequences. *Bioinformatics* **2018**, *34*, 237–244. [[CrossRef](#)] [[PubMed](#)]
95. Jurtz, V.I.; Rosenberg Johansen, A.; Nielsen, M.; Almagro Armenteros, J.J.; Nielsen, H.; KAAe Sønderby, C.; Winther, O.; Kaae Sønderby, S. An introduction to deep learning on biological sequence data: Examples and solutions. *Bioinformatics* **2017**, *33*, 3685–3690. [[CrossRef](#)] [[PubMed](#)]
96. Yang, Z.R.; Yang, Z. 6.01—Artificial Neural Networks. In *Comprehensive Biomedical Physics*; Brahme, A., Ed.; Elsevier: Amsterdam, The Netherlands, 2014. [[CrossRef](#)]
97. Khosrow-Pour, M.; Johnston, L.; Yoder, J. Analysis of Large-Scale OMIC Data Using Self Organizing Maps. In *Encyclopedia of Information Science and Technology*; Khosrow-Pour, M., Ed.; IGI Global: Hershey, PA, USA, 2014.
98. You, Y.; Lai, X.; Pan, Y.; Zheng, H.; Vera, J.; Liu, S.; Deng, S.; Zhang, L. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct. Targeted Ther.* **2022**, *7*, 156. [[CrossRef](#)]
99. Wei, M.; Zhang, X.; Pan, X.; Wang, B.; Ji, C.; Qi, Y.; Zhang, J.Z.H. HobPre: Accurate prediction of human oral bioavailability for small molecules. *J. Cheminf.* **2022**, *14*, 1. [[CrossRef](#)]
100. Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. Reinvent 2.0: An AI tool for de novo drug design. *J. Chem. Inf. Model.* **2020**, *60*, 5918–5922. [[CrossRef](#)] [[PubMed](#)]
101. Chen, S.; Li, Z.; Zhang, S.; Zhou, Y.; Xiao, X.; Cui, P.; Xu, B.; Zhao, Q.; Kong, S.; Dai, Y. Emerging biotechnology applications in natural product and synthetic pharmaceutical analyses. *Acta Pharm. Sin. B* **2022**, *12*, 4075–4097. [[CrossRef](#)] [[PubMed](#)]
102. Rim, K.T. In silico prediction of toxicity and its applications for chemicals at work. *Toxicol. Environ. Health Sci.* **2020**, *12*, 191–202. [[CrossRef](#)]
103. Yang, H.; Sun, L.; Li, W.; Liu, G.; Tang, Y. In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front. Chem.* **2018**, *6*, 30. [[CrossRef](#)]
104. Panapitiya, G.; Girard, M.; Hollas, A.; Sepulveda, J.; Murugesan, V.; Wang, W.; Saldanha, E. Evaluation of deep learning architectures for aqueous solubility prediction. *ACS Omega* **2022**, *7*, 15695–15710. [[CrossRef](#)]
105. Ye, Z.; Ouyang, D. Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. *J. Cheminf.* **2021**, *13*, 98. [[CrossRef](#)]
106. Agüero-Chapin, G.; Galpert-Canizares, D.; Dominguez-Perez, D.; Marrero-Ponce, Y.; Perez-Machado, G.; Teijeira, M.; Antunes, A. Emerging computational approaches for antimicrobial peptide discovery. *Antibiotics* **2022**, *11*, 936. [[CrossRef](#)]
107. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803. [[CrossRef](#)]
108. Covell, D.G.; Huang, R.; Wallqvist, A. Anticancer medicines in development assessment of bioactivity profiles within the National Cancer Institute anticancer screening data. *Mol. Cancer Therapeut.* **2007**, *6*, 2261–2270. [[CrossRef](#)]

109. Jan, A.; Hayat, M.; Wedyan, M.; Alturki, R.; Gazzawe, F.; Ali, H.; Alarfaj, F.K. Target-AMP: Computational prediction of antimicrobial peptides by coupling sequential information with evolutionary profile. *Comput. Biol. Med.* **2022**, *151 Pt A*, 106311. [[CrossRef](#)]
110. Asma, S.T.; Acaroz, U.; Imre, K.; Morar, A.; Shah, S.R.A.; Hussain, S.Z.; Arslan-Acaroz, D.; Demirbas, H.; Hajrulai-Musliu, Z.; Istanbulgul, F.R.; et al. Natural Products/Bioactive Compounds as a Source of Anticancer Drugs. *Cancers* **2022**, *14*, 6203. [[CrossRef](#)]
111. Venkatesh, G.; Grover, A.; Srinivasaraghavan, G.; Rao, S. MHCAttnNet: Predicting MHC-peptide bindings for MHC alleles classes I and II using an attention-based deep neural model. *Bioinformatics* **2020**, *36*, 399–406. [[CrossRef](#)]
112. Jiang, L.; Tang, J.; Guo, F.; Guo, Y. Prediction of Major Histocompatibility Complex Binding with Bilateral and Variable Long Short Term Memory Networks. *Biology* **2022**, *11*, 848. [[CrossRef](#)]
113. Yang, A.; Zhang, W.; Wang, J.; Yang, K.; Han, Y.; Zhang, L. Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1032. [[CrossRef](#)]
114. Bellazzi, R.; Diomidous, M.; Sarkar, N.; Takabayashi, K.; Ziegler, A.; McCray, A.T. Data Analysis and Data Mining: Current Issues in Biomedical Informatics. *Methods Inf. Med.* **2011**, *50*, 536–544. [[CrossRef](#)]
115. Zameer, R.; Tariq, S.; Noreen, S.; Sadaqat, M.; Azeem, F. Role of Transcriptomics and Artificial Intelligence Approaches for the Selection of Bioactive Compounds, Chapter 10. In *Drug Design Using Machine Learning*; Inamuddin, Altalhi, T., Jorddy, N., Cruz, N., Refat, M.S.E.-D., Eds.; Wiley: Hoboken, NJ, USA, 2022. [[CrossRef](#)]
116. Gabere, M.N.; Noble, W.S. Empirical comparison of web-based antimicrobial peptide prediction tools. *Bioinformatics* **2017**, *33*, 1921–1929. [[CrossRef](#)]
117. Thomas, S.; Karnik, S.; Barai, R.S.; Jayaraman, V.K.; Idicula-Thomas, S. CAMP: A useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* **2010**, *38*, gkp1021. [[CrossRef](#)]
118. Wagh, F.H.; Barai, R.S.; Gurung, P.; Idicula-Thomas, S. CAMP<sub>R3</sub>: A database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* **2016**, *44*, D1094–D1097. [[CrossRef](#)]
119. Lawrence, T.J.; Carper, D.L.; Spangler, M.K.; Carrell, A.A.; Rush, T.A.; Minter, S.J.; Weston, D.J.; Labbé, J.L. amPEPpy 1.0: A portable and accurate antimicrobial peptide prediction tool. *Bioinformatics* **2021**, *37*, 2058–2060. [[CrossRef](#)]
120. Esteban, A.; Giraldo, G.P.; Orduz, S. InverPep: A database of invertebrate antimicrobial peptides. *J. Glob. Antimicrob. Resist.* **2017**, *8*, 13–17. [[CrossRef](#)]
121. Lata, S.; Mishra, N.K.; Raghava, G.P. AntiBP2: Improved version of antibacterial peptide prediction. *BMC Bioinform.* **2010**, *11*, S19. [[CrossRef](#)]
122. Torrent, M.; Di Tommaso, P.; Pulido, D.; Nogués, V.; Notredame, C.; Boix, E.; Andreu, D. AMPA: An automated web server for prediction of protein antimicrobial regions. *Bioinformatics* **2012**, *28*, 130–131. [[CrossRef](#)]
123. Tyagi, A.; Kapoor, P.; Kumar, P.; Chaudhary, K.; Gautam, A.; Raghava, G.P.S. In Silico Models for Designing and Discovering Novel Anticancer Peptides. *Sci. Rep.* **2013**, *3*, 2984. [[CrossRef](#)]
124. Phan, L.T.; Park, H.W.; Pitti, T.; Madhavan, T.; Jeon, Y.-J.; Manavalan, B. MLACP 2.0: An updated machine learning tool for anticancer peptide prediction. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4473–4480. [[CrossRef](#)]
125. Wu, X.; Zeng, W.; Lin, F.; Xu, P.; Li, X. Anticancer Peptide Prediction via Multi-Kernel CNN and Attention. *Model. Front. Genet.* **2022**, *13*, 887894. [[CrossRef](#)]
126. Chen, W.; Ding, H.; Feng, P.; Lin, H.; Chou, K.-C. iACP: A Sequence-Based Tool for Identifying Anticancer Peptides. *Oncotarget* **2016**, *7*, 16895–16909. [[CrossRef](#)]
127. Vijayakumar, S.; Ptv, L. ACP: A Web Server for Prediction and Design of Anti-cancer Peptides. *Int. J. Pept. Res. Ther.* **2015**, *21*, 99–106. [[CrossRef](#)]
128. Akbar, S.; Hayat, M.; Iqbal, M.; Jan, M.A. iACP-GAEnsC: Evolutionary genetic algorithm, based ensemble classification of anticancer peptides by utilizing hybrid feature space. *Artif. Intell. Med.* **2017**, *79*, 62–70. [[CrossRef](#)]
129. Kabir, M.; Arif, M.; Ahmad, S.; Ali, Z.; Swati, Z.N.K.; Yu, D.-J. Intelligent computational method for discrimination of anticancer peptides by incorporating sequential and evolutionary profiles information. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 158–165. [[CrossRef](#)]
130. Schaduangrat, N.; Nantasenamat, C.; Prachayasittikul, V.; Shoombuatong, W. ACPred: A Computational Tool for the Prediction and Analysis of Anticancer Peptides. *Molecules* **2019**, *24*, 1973. [[CrossRef](#)]
131. Wei, L.; Zhou, C.; Chen, H.; Song, J.; Su, R. ACPred-FL: A sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* **2018**, *34*, 4007–4016. [[CrossRef](#)]
132. Rao, B.; Zhou, C.; Zhang, G.; Su, R.; Wei, L. ACPred-Fuse: Fusing multi-view information improves the prediction of anticancer peptides. *Brief. Bioinform.* **2020**, *21*, 1846–1855. [[CrossRef](#)]
133. Yi, H.-C.; You, Z.-H.; Zhou, X.; Cheng, L.; Li, X.; Jiang, T.-H.; Chen, Z.-H. ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High-Efficiency Feature Representation. *Mol. Ther.-Nucleic Acids* **2019**, *17*, 1–9. [[CrossRef](#)]
134. Charoenkwan, P.; Chiangjong, W.; Lee, V.S.; Nantasenamat, C.; Hasan, M.M.; Shoombuatong, W. Improved prediction and characterization of anticancer activities of peptides using a novel flexible scoring card method. *Sci. Rep.* **2021**, *11*, 3017. [[CrossRef](#)]
135. Andreatta, M.; Trolle, T.; Yan, Z.; Greenbaum, J.A.; Peters, B.; Nielsen, M. An automated benchmarking platform for MHC class II binding prediction methods. *Bioinformatics* **2018**, *34*, 1522–1528. [[CrossRef](#)]

136. Vita, R.; Mahajan, S.; Overton, J.A.; Dhanda, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2018**, *47*, D339–D343. [[CrossRef](#)] [[PubMed](#)]
137. Jensen, K.K.; Andreatta, M.; Marcatili, P.; Buus, S.; Greenbaum, J.A.; Yan, Z.; Sette, A.; Peters, B.; Nielsen, M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **2018**, *154*, 394–406. [[CrossRef](#)] [[PubMed](#)]
138. Bui, H.H.; Sidney, J.; Peters, B.; Sathiamurthy, M.; Sinichi, A.; Purton, K.A.; Mothé, B.R.; Chisari, F.V.; Watkins, D.I.; Sette, A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* **2005**, *57*, 304–314. [[CrossRef](#)] [[PubMed](#)]
139. Doytchinova, I.A.; Flower, D.R. Towards the in-silico identification of class II restricted T-cell epitopes: A partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics* **2003**, *19*, 2263–2270. [[CrossRef](#)] [[PubMed](#)]
140. Sturniolo, T.; Bono, E.; Ding, J.; Raddrizzani, L.; Tuereci, O.; Sahin, U.; Braxenthaler, M.; Gallazzi, F.; Protti, M.P.; Sinigaglia, F.; et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* **1999**, *17*, 555–561. [[CrossRef](#)] [[PubMed](#)]
141. Dhanda, S.K.; Usmani, S.S.; Agrawal, P.; Nagpal, G.; Gautam, A.; Raghava, G.P.S. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief. Bioinform.* **2016**, *18*, 467–478.
142. Armenteros, A.J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)] [[PubMed](#)]
143. Grønning, A.G.B.; Kacprowski, T.; Schéele, C. MultiPep: A hierarchical deep learning approach for multi-label classification of peptide bioactivities. *Biol. Methods Protoc.* **2021**, *6*, bpab021. [[CrossRef](#)]
144. Mooney, C.; Haslam, N.J.; Pollastri, G.; Shield, D.C. Towards the Improved Discovery and Design of Functional Peptides: Common Features of Diverse Classes Permit Generalized Prediction of Bioactivity. *PLoS ONE* **2012**, *7*, e0045012. [[CrossRef](#)]
145. Mooney, C.; Haslam, N.J.; Thérèse, A.; Pollastri, G.; Shields, D.C. PeptideLocator: Prediction of bioactive peptides in protein sequences. *Bioinformatics* **2013**, *29*, 1120–1126. [[CrossRef](#)]
146. Veltri, D.; Kamath, U.; Shehu, A. Deep learning improves antimicrobial peptide recognition. *Bioinformatics* **2018**, *34*, 2740–2747. [[CrossRef](#)] [[PubMed](#)]
147. Yan, J.; Bhadra, P.; Li, A.; Sethiya, P.; Qin, L.; Tai, H.K.; Wong, K.H.; Siu, S. Deep-AmPEP30: Improve short antimicrobial peptides prediction with deep learning. *Mol. Ther. Nucleic Acids* **2020**, *20*, 882–894. [[CrossRef](#)] [[PubMed](#)]
148. Hussain, W. sAMP-PFPDeep: Improving accuracy of short antimicrobial peptides prediction using three different sequence encodings and deep neural networks. *Brief. Bioinform.* **2022**, *23*, bbab487. [[CrossRef](#)] [[PubMed](#)]
149. Van Oort, C.M.; Ferrell, J.B.; Remington, J.M.; Wshah, S.; Li, J. AMPGAN v2: Machine Learning-Guided Design of Antimicrobial Peptides. *Chem. Inf. Model.* **2021**, *61*, 2198–2207. [[CrossRef](#)]
150. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **2019**, *47*, W402–W407. [[CrossRef](#)]
151. Zhang, C.; Shine, M.; Pyle, A.M.; Zhang, Y. US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **2022**, *19*, 1109–1115. [[CrossRef](#)]
152. Raghava, G.P.S. APSSP2: A Combination Method for Protein Secondary Structure Prediction Based on Neural Network and Example-Based Learning. CASP5. A-132, Which Using Nearest Neighbor and Neural Network Approach to Perform Predictions. 2002. Available online: <http://crdd.osdd.net/raghava/apssp2/> (accessed on 20 October 2023).
153. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
154. McGuffin, L.J.; Edmunds, N.S.; Genc, A.G.; Alharbi, S.M.A.; Salehe, B.R.; Adiyaman, R. Prediction of protein structures, functions and interactions using the IntFOLD7, MultiFOLD and ModFOLDdock servers. *Nucleic Acids Res.* **2023**, *51*, W274–W280. [[CrossRef](#)]
155. Morten, K.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **2012**, *7*, 1511–1522.
156. Lambert, C.; Leonard, N.; De Bolle, X.; Depiereux, E. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* **2002**, *18*, 1250–1256. [[CrossRef](#)]
157. Krobthong, S.; Yingchutrakul, Y.; Wongtrakoongate, P.; Chuntakaruk, H.; Rungrotmongkol, T.; Chaichana, C.; Mahatnirunkul, T.; Chomtong, T.; Choowongkamon, K.; Aonbangkhen, C. Proteomics and Molecular Docking Analyses Reveal the BioChemical and Molecular Mechanism Underlying the Hypolipidemic Activity of Nano-Liposomal Bioactive Peptides in 3T3-L1 Adipocytes. *Foods* **2023**, *12*, 780. [[CrossRef](#)] [[PubMed](#)]
158. Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157. [[CrossRef](#)] [[PubMed](#)]
159. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C.R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7–S26. [[CrossRef](#)] [[PubMed](#)]
160. McConkey, B.J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Sci.* **2002**, *83*, 845–855.

161. Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V.V.; Tanchuk, V.Y.; et al. Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* **2011**, *25*, 533–554. [[CrossRef](#)]
162. Dong, J.; Yao, Z.J.; Zhu, M.F.; Wang, N.N.; Lu, B.; Chen, A.F.; Lu, A.P.; Miao, H.; Zeng, W.B.; Cao, D.S. ChemSAR: An online pipelining platform for molecular SAR modeling. *J. Cheminform.* **2017**, *9*, 27. [[CrossRef](#)]

**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.