


## Review

# Conotoxins: Classification, Prediction, and Future Directions in Bioinformatics

Rui Li, Junwen Yu, Dongxin Ye, Shanghua Liu, Hongqi Zhang, Hao Lin , Juan Feng and Kejun Deng 

The Clinical Hospital of Chengdu Brain Science Institute, School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China; 202321140412@std.uestc.edu.cn (R.L.); 202322140206@std.uestc.edu.cn (J.Y.); 202321140113@std.uestc.edu.cn (D.Y.); 202321140418@std.uestc.edu.cn (S.L.); 202321140413@std.uestc.edu.cn (H.Z.); hlin@uestc.edu.cn (H.L.); fengjuan@uestc.edu.cn (J.F.)

\* Correspondence: dengkj@uestc.edu.cn

**Abstract:** Conotoxins, a diverse family of disulfide-rich peptides derived from the venom of *Conus* species, have gained prominence in biomedical research due to their highly specific interactions with ion channels, receptors, and neurotransmitter systems. Their pharmacological properties make them valuable molecular tools and promising candidates for therapeutic development. However, traditional conotoxin classification and functional characterization remain labor-intensive, necessitating the increasing adoption of computational approaches. In particular, machine learning (ML) techniques have facilitated advancements in sequence-based classification, functional prediction, and de novo peptide design. This review explores recent progress in applying ML and deep learning (DL) to conotoxin research, comparing key databases, feature extraction techniques, and classification models. Additionally, we discuss future research directions, emphasizing the integration of multimodal data and the refinement of predictive frameworks to enhance therapeutic discovery.

**Keywords:** conotoxins; venom peptide identification; computational peptide analysis; machine learning applications; peptide-based drug discovery

**Key Contribution:** This work provides a comprehensive review of conotoxin bioinformatics, emphasizing advancements in sequence classification, database development, and the use of machine learning and deep learning for predicting conotoxin function and receptor interactions, offering insights for future drug discovery and targeted therapies.



Received: 28 December 2024

Revised: 25 January 2025

Accepted: 7 February 2025

Published: 9 February 2025

**Citation:** Li, R.; Yu, J.; Ye, D.; Liu, S.; Zhang, H.; Lin, H.; Feng, J.; Deng, K. Conotoxins: Classification, Prediction, and Future Directions in Bioinformatics. *Toxins* **2025**, *17*, 78. <https://doi.org/10.3390/toxins17020078>

**Copyright:** © 2025 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

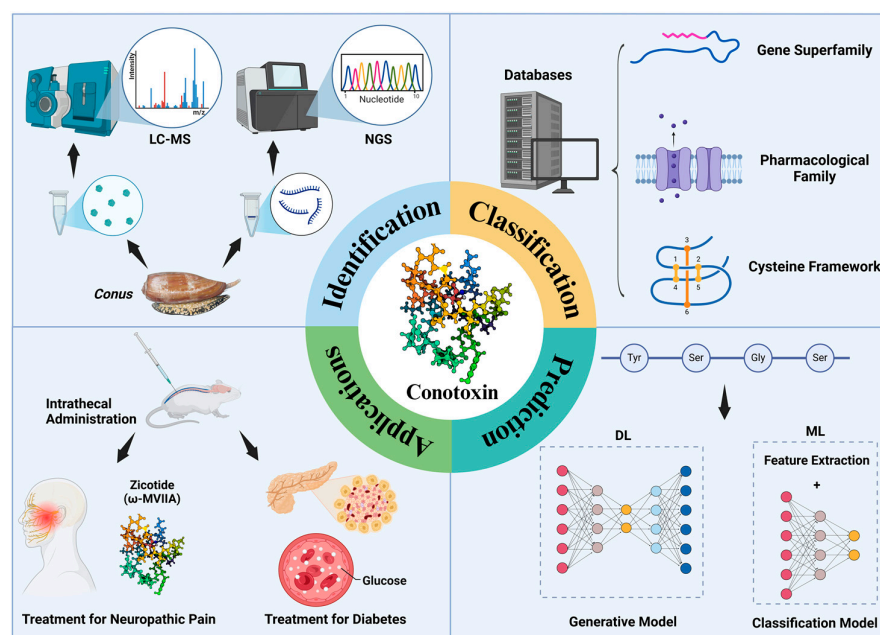
Conotoxins are small peptides found in the venom of cone snails (genus *Conus*), and they have attracted significant attention due to their potent pharmacological effects, particularly in modulating ion channels, receptors, and neurotransmitter systems [1]. They are important in regulating key biological processes, such as nerve signal transmission, making them valuable in fundamental and applied biomedical research. Given their high specificity and diverse mechanisms of action, conotoxins hold immense promise as therapeutic agents, particularly in pain management, neurological disorders, and other diseases with inadequate treatment options [2].

Despite their significant potential, the discovery, characterization, and therapeutic optimization of conotoxins face substantial challenges [3–6]. Traditionally, conotoxin discovery relies on extracting and isolating active components from *Conus* venom [7–9], followed by classification into various superfamilies and pharmacological families based on

sequence homology and functional properties. However, traditional methods for analysis and identification are time-consuming and labor-intensive [10]. As a result, researchers have increasingly turned to computational approaches, particularly bioinformatics tools and machine learning (ML) techniques [11–14], to improve the classification, prediction, and generation efficiency of conotoxins [15–27].

Integrating large-scale databases, such as UniProt, ConoServer, and other specialized repositories, has significantly accelerated conotoxin research [28–31]. These resources provide comprehensive sequence, structure, and functional information, supporting the development of predictive models and facilitating the discovery of novel toxins. In recent years, machine learning and deep learning (DL) methods have proven highly effective [32–35], enabling researchers to predict the functions of conotoxins, identify potential receptor targets, and even generate new peptide sequences for drug development [36–38]. There remains considerable room for expanding the data on conotoxins. The complexity of their sequences, the broad diversity of their targets, and the inherent structural diversity of conotoxins present challenges for constructing more diversified and high-accuracy classification and prediction models [39,40]. Additionally, the integration of multi-modal data, such as structural information and post-translational modifications, remains underexplored. These gaps highlight the need for more advanced computational approaches and comprehensive databases to accelerate conotoxin discovery and application.

This review introduces the properties, classification, and applications of conotoxins, highlighting research in bioinformatics, including conotoxin databases, predictive models, and deep learning-based generation frameworks (Figure 1). It also provides an outlook on future research directions, such as database optimization, multi-modal data integration, and advanced modeling approaches.



**Figure 1.** A schematic diagram of the content covered in this review. Conotoxin research encompasses four major aspects: (1) Identification, which has progressed from traditional chemical methods (e.g., MS) to high-throughput genomic and transcriptomic technologies, significantly improving discovery efficiency. (2) Classification, where conotoxins are grouped based on evolutionary superfamilies, pharmacological families, and cysteine frameworks. (3) Application, highlighting their therapeutic potential, such as intrathecal drug development for pain relief and diabetes treatment. (4) Prediction, where computational approaches, including machine learning, facilitate conotoxin sequence generation and classification. This figure summarizes the transition from traditional experimental techniques to modern computational bioinformatics in conotoxin research.

## 2. Conotoxins

### 2.1. Sequence Characteristics of Conotoxins

Conotoxins are a group of neurotoxic peptides isolated from the venom of *Conus* species, characterized by 10 to 40 amino acid residues and a high content of disulfide bonds, exhibiting high selectivity and biological activity [5]. They serve as useful high-affinity ligands for various receptors and ion channels [41], including ion channels, G protein-coupled receptors (GPCRs), transporters, and enzymes [42,43]. One of the most notable features of conotoxins is the high abundance of cysteine residues in their sequences, which form disulfide bonds that stabilize the three-dimensional structure of the toxin and confer resistance to degradation [44].

The genes encoding bioactive venom components are characterized by accelerated evolution, and conotoxins may represent one of the fastest-evolving gene products known to date [10,45,46]. As a precursor for conotoxin biosynthesis, its transcript sequences consist of three regions: the endoplasmic reticulum (ER) signal peptide, the mature peptide region, and pre- or post-mature peptide region [41]. These regions are crucial in determining the toxin's synthesis, functionality, and targeting properties [5].

Conotoxin sequences exhibit significant diversity across different species [47], with considerable variation in the order of the mature peptide region. Only a small subset of conotoxins are expressed in two or more *Conus* species [30]. In contrast, the ER signal peptide is highly conserved within specific subgroups and is typically used to cluster different conotoxins into corresponding gene superfamilies [42,48]. These conserved amino acid regions are directly associated with the biological activity of the toxin. They exert their effects by binding to ion channels in the nervous system, such as sodium, calcium, and potassium channels, or receptors, thereby inhibiting neurotransmission or modulating neural signaling [5].

Conotoxins typically undergo various post-translational modifications, many of which cannot be predicted from the precursor sequence. The formation of disulfide bonds is the most common post-translational modification in conotoxins [49]. In addition, most conotoxins carry additional post-translational modifications, such as hydroxylation of proline [49], sulfonation of tyrosine [50], and hydroxylation of glutamic acid [51]. These modifications further enhance the stability and biological activity of the toxins [52].

The structural and functional complexity of conotoxins is largely influenced by their post-translational modifications, which play a crucial role in defining their stability, bioactivity, and target specificity [50]. The formation of disulfide bonds, the most prevalent modification, is essential for maintaining the three-dimensional conformation of these peptides, thereby enhancing their resistance to enzymatic degradation and ensuring their biological function [53]. Additionally, modifications such as hydroxylation of proline, sulfonation of tyrosine, and hydroxylation of glutamic acid further refine their molecular interactions, enabling conotoxins to exhibit remarkable selectivity toward ion channels and receptors. These biochemical alterations not only contribute to their pharmacological potency but also add another layer of complexity to their classification and functional annotation [54,55]. With the advancement of high-throughput sequencing and bioinformatics approaches, researchers are now better equipped to explore the impact of modifications on conotoxin diversity and activity, enhancing our understanding of their evolutionary significance and potential applications in biomedical research. For example, next-generation sequencing (NGS) technologies like Illumina and PacBio sequencing enable the rapid sequencing of venom gland transcriptomes, revealing novel conotoxins and their associated modifications [10,56]. Transcriptome sequencing was used to uncover new conotoxins in *Conus betulinus*, demonstrating the power of these methods [57]. Bioinformatics tools such as ConoServer and ConoMode assist in annotating and classifying conotoxins [28–30,58],

while molecular dynamics simulations (e.g., using GROMACS or AMBER) provide insights into how structural features influence conotoxin activity [59,60]. These technologies and approaches are essential for advancing conotoxin research and expanding their biomedical potential.

## 2.2. Discovery and Identification of Conotoxins

Since their initial discovery in the 1960s [61], research on conotoxins has evolved from traditional chemical analysis to modern genomics and bioinformatics, which significantly enhanced the efficiency and accuracy of their discovery and identification ability.

Initially, the discovery of conotoxins relied primarily on the extraction and isolation technology from *Conus* venom. Researchers isolated various peptide molecules through chemical analysis techniques such as high-performance liquid chromatography (HPLC) and mass spectrometry (MS) [7,9] and verified their toxicity through bioactivity assays, including cell-based cytotoxicity assays, enzyme inhibition assays, and receptor binding assays [62,63]. However, as the diversity and functional complexity of conotoxins increased, traditional chemical separation methods began to face limitations in terms of time and sample volume, making it difficult to meet the demands for high-throughput screening of large numbers of toxins.

To address this challenge, modern discovery and identification methods have shifted toward high-throughput genomic and transcriptomic technologies [64–66]. Advanced sequencing techniques and bioinformatics tools reduced the need to separate toxins from venom. They can help to reveal tens of thousands of previously unknown animal venom peptides and protein sequences, including over 20,000 conotoxin sequences [6,67]. As of 30 November 2024, the search result for “venom duct transcriptome” provided transcriptomic data from various species and taxonomic groups. There was a total of 126 entries related to venom duct transcriptomes. These entries represent a range of biological groups, including chordates, arthropods, and mollusks. Excluding the seven entries from the species *Raphitoma purpurea* (family Raphitomidae), all other entries originate from the family Conidae, covering 19 species within the genus *Conus*. In addition, a retrospective review by Jin et al. in 2019 documented the raw reads, accession numbers, and references for the venom gland transcriptomes of 30 *Conus* species [5].

These data provide valuable resources for marine venom gland functional studies, enabling researchers to perform further comparative analyses and functional investigations. By sequencing the genomes and transcriptomes of conotoxins, researchers can directly predict and identify toxin genes at the genetic level. This approach allows for the faster discovery of novel toxins and expands the range of toxins identified [3,68]. Compared to traditional venom extraction methods, genomic and transcriptomic-based discovery can significantly increase efficiency and offer strong technical support for systematically studying the diversity of conotoxins.

Currently, many studies utilize a combination of proteomics and transcriptomics to explore conotoxins in the venom of individual *Conus* species [69–71]. Next-generation sequencing (NGS) of venom gland transcriptomes provides an unbiased list of precursor sequences [5,42]. In addition to traditional proteomic techniques, advanced methods such as X-ray crystallography [72] and nuclear magnetic resonance NOESY [73,74]/ROESY [75] experiments are used to identify mature peptides and post-translational modifications.

The methods for identifying conotoxins have undergone a revolutionary transformation applying bioinformatics and computational biology techniques. Traditional identification methods rely on the separation and purification of venom, which often face challenges such as difficulties in sample acquisition (especially for rare and/or dangerous species) and are time-consuming [10]. Meanwhile, modern identification methods

integrated high-throughput gene sequencing and bioinformatics tools, allowing for the direct prediction of toxins from genes and transcripts based on sequence information. Sequence alignment tools such as BLAST [76] and HMMER [77] have been used for identifying toxin genes. Compared with known toxins, these tools can quickly identify and validate new toxins [3,15,71,78]. Machine learning methods are widely employed in the identification and classification of proteins or peptides, considering features such as amino acid composition, n-mer amino acid composition, pseudo-amino acid composition, and position-specific scoring matrices (PSSM) [15–21]. These computational approaches not only provide new ways to rapidly and accurately identify conotoxins and their categories but also offer possibilities for a deeper understanding of their functional characteristics, overcoming the limitations of traditional biochemical experimental methods [79].

In conclusion, the discovery and identification of conotoxins have evolved from traditional chemical analysis to a combination of genomics, transcriptomics, and computational biology techniques. Modern genomic and bioinformatics methods have greatly enhanced the efficiency of toxin discovery and identification, providing support for the functional prediction, classification, and drug development of novel toxins.

### 2.3. Classification of Conotoxins

As research on conotoxins has advanced, these toxins have been classified into different cysteine frameworks, gene superfamilies, or pharmacological families based on factors such as the cysteine motifs in their mature peptide regions, the similarity of the precursor endoplasmic reticulum (ER) signal sequences, and their specificity for pharmacological targets [28,29,80]. The classification of conotoxins was essential for understanding their biological functions and provided a theoretical basis for drug development and neuroscience research.

#### 2.3.1. Cysteine Patterns of Conotoxins

A distinctive feature of conotoxins is their high content of cysteine residues. During the maturation process of proteins in eukaryotes, cysteine residues were oxidized to form disulfide bonds, which stabilize the three-dimensional structure of the protein [81,82]. The arrangement and number of disulfide bonds vary among toxins from different superfamilies and families, and these differences not only influence the stability of the toxin but also have significant effects on its binding affinity to targets and its toxicity [83]. The diversity in the patterns of disulfide bond connectivity allows for the classification of conotoxins based on their cysteine frameworks [80]. A cysteine framework was defined by the number of cysteine residues and the number of residues (either none or at least one) between adjacent cysteine residues [28,29]. There were 30 framework families classified in ConoServer based on the number of cysteine residues, their ring sizes, and the different disulfide bond linkages. A significant portion of these cysteine frameworks has been reported in transcriptomic studies.

#### 2.3.2. Superfamilies of Conotoxins

The classification of conotoxin superfamilies is based on the similarity of their signal peptide sequences, the arrangement of disulfide bonds, and their evolutionary relationships. Each superfamily encompasses a group of toxins that share conserved regions but exhibit functional diversity [5]. These toxins possess similar structures, and conserved signal peptide sequences. They target a wide range of receptors or ion channels within the nervous system.

As of the current understanding, the identified superfamilies of conotoxins have been divided into over 20 major categories, which can further subdivide into different branches based on the amino acid characteristics of their frameworks. Moreover, some



conotoxins that have been characterized do not belong to any of the established gene superfamilies [80,84,85] and are classified into temporary gene superfamilies in the ConoServer database [28]. Well-known superfamilies, including A, M, P, O, S, T, and I, were associated with distinct biological functions and mechanisms of action. For instance, conotoxins from the A superfamily predominantly feature a type I cysteine framework (CC-C-C) and are known to effectively and selectively target a range of neuronal and neuromuscular nAChR subtypes [80]. Conantokin-G, from the B superfamily, is a non-competitive N-methyl-D-aspartate receptor (NMDAR) antagonist [86] and is currently under extensive investigation as a potential drug for pain relief, anticonvulsant therapy, and treatment of other neurological conditions [87–89]. The M superfamily, including branches M-4 and M-5, targets molecular sites such as voltage-gated potassium channels, acetylcholine receptors, and voltage-gated sodium channels, which can be further categorized into distinct pharmacological families [90,91].

In addition to the well-established superfamilies mentioned above, Robinson et al. also recorded four additional superfamilies not listed in the ConoServer database, namely ConoCAPs, Conopressins/Conophysins, Konkunitzins, and Con-Ikot-Ikots, as well as some conotoxins that have not yet been classified [80]. Table 1 summarizes the superfamilies mentioned above, along with their associated cysteine frameworks and pharmacological families. Scientists can further categorize these toxins into superfamilies or families utilizing machine learning methods in combination with sequence alignment and evolutionary analysis, thereby revealing their biological characteristics.

**Table 1.** Classification of gene superfamilies and their associated cysteine frameworks and pharmacological families. The table includes only the established gene superfamilies and does not include temporary gene superfamilies. Data sourced from ConoServer and [80].

Gene Superfamily	Cysteine Framework	Pharmacological Family
A	I, II, IV, VI/VII, XIV, XXII	$\alpha$ , $\kappa$ , $\rho$ , other
B1		
B2	VIII	
B3	XXIV	$\alpha$
C		$\alpha$
D	IV, XIV, XV, XX, XXIV, XXVIII	$\alpha$
E	XXII	
F		
G	XIII	
H	VI/VII	
I1	VI/VII, XI, XXII	$\iota$
I2	VI/VII, XI, XII, XIII, XIV	$\kappa$
I3	VI/VII, XI	
J	XIV	$\alpha$ , $\kappa$
K	XXIII	
L	XIV, XXIV	$\alpha$
M	I, II, III, IV, VI/VII, IX, XIV, XVI, XXXII	$\alpha$ , $\iota$ , $\kappa$ , $\mu$
N	XV	
O1	I, VI/VII, IX, XII, XIV, XVI, XXIX	$\delta$ , $\gamma$ , $\kappa$ , $\mu$ , $\omega$
O2	I, VI/VII, XII, XIV, XV, XVI	$\gamma$
O3	VI/VII, XVI	
P	IX, XIV	

Table 1. Cont.

Gene Superfamily	Cysteine Framework	Pharmacological Family
Q	VI/VII, XVI	$\alpha$ , $\sigma$
R	XIV	
S	VIII, XXXIII	
T	I, V, X, XVI	
U	VI/VII	
V	XV	
Y	VI/VII, XVII	
conodipine	XXI,	$\kappa$
con-ikot-ikot		
conoCAP		
conopressin		
conkunitzin	XIV	

### 2.3.3. Pharmacological Families of Conotoxins

Conotoxins are classified into numerous pharmacological families based on the type of their molecular targets and their corresponding pharmacological activities [1]. These toxins exhibit highly diverse structures and functions but share a common specificity for receptors. The primary targets of conotoxins are membrane proteins, particularly ion channels, membrane receptors, and transporters [92]. Conotoxins exert their biological effects by binding to various receptors or ion channels in the nervous system, modulating the transmission of neural signals and neuronal excitability [5]. Due to their ability to selectively interact with specific pharmacological targets, conotoxins are widely used as pharmacological tools in research related to disease treatment [93,94].

Conotoxins are widely used to study ion channel functions and their dysfunctions [67]. Ion channel-targeting conotoxins interact with specific ion channels, altering their permeability and regulating neuronal electrical activity. Based on the type of ion channel targeted, ion channel-specific conotoxins can be further categorized as follows:

- (1) Sodium Channel-Targeting Toxins (Na-Conotoxins):  $\mu$ -conotoxins act as classical pore blockers, binding directly to the sodium channel pore and inhibiting sodium ion flux [90,95].  $\delta$ -conotoxins and  $\iota$ -conotoxins, on the other hand, function as gating modifiers, modulating the gating states of sodium channels, thereby affecting sodium ion flow and regulating the propagation of neural signals [53,96].
- (2) Calcium Channel-Targeting Toxins (Ca-Conotoxins):  $\omega$ -conotoxins block the passage of  $\text{Ca}^{2+}$  through voltage-gated calcium channels (CaVs) at presynaptic terminals, thereby disrupting the release of acetylcholine-containing vesicles and interfering with neurotransmission [97].  $\omega$ -MVIIA conotoxin is the only conotoxin approved by the FDA for the treatment of severe chronic pain [98–100].
- (3) Potassium Channel-Targeting Toxins (K-Conotoxins):  $\mu$ -conotoxin, although primarily known as a sodium voltage-gated channel inhibitor, also exhibits similar blocking effects on Shaker K channels [101] and mammalian Kv1.2 channels [102]. This result suggests that certain toxins can exert broad effects on different types of ion channels.

In addition to ion channels, some conotoxins can target neurotransmitter receptors and other membrane proteins [43]. These pharmacological families are documented in Table 2. For example, the  $\alpha$ -conotoxin family interacts with nicotinic acetylcholine receptors (nAChRs) to modulate neurotransmitter release and neuronal plasticity, potentially offering neuroprotective effects [103].  $\rho$ -conotoxins may regulate G-protein coupled receptor (GPCR) signaling through interaction with  $\alpha_1$ -adrenergic receptors ( $\alpha_1$ -AR) [104]. Members of the neurohypophyseal conotoxin family, such as contulakin-G (CGX-1160), act on GPCRs and play critical roles in neurotransmission and neuromodulation [105].

The diverse mechanisms of action of these toxins highlight their significant potential in neuropharmacological research and clinical applications, including the treatment of pain, epilepsy, anesthesia, and neuroinjury [106]. Classification based on their targets facilitates a deeper understanding of the biological functions of these toxins and provides a theoretical foundation for the development of targeted therapeutics [79].

**Table 2.** Pharmacological families used in ConoServer. This table summarizes the pharmacological families of conotoxins as classified in ConoServer. Each family is associated with specific molecular targets, including ion channels and receptors, and their representative toxins are listed for reference.

Family	Definition	Representative Toxin(s)
$\alpha$ (alpha)	Nicotinic acetylcholine receptors (nAChR)	GI
$\gamma$ (gamma)	Neuronal pacemaker cation currents (inward cation current)	PnVIIA, TxVIIA
$\delta$ (delta)	Voltage-gated Na channels (agonist, delay inactivation)	TxVIA
$\epsilon$ (epsilon)	Presynaptic Ca channels or G protein-coupled presynaptic receptors	TxVA
$\iota$ (iota)	Voltage-gated Na channels (agonist, no delayed inactivation)	RXIA
$\kappa$ (kappa)	Voltage-gated K channels (blocker)	PVIIA
$\mu$ (mu)	Voltage-gated Na channels (antagonist, blocker)	GIIIA
$\rho$ (rho)	Alpha1-adrenoceptors (GPCR)	TIA
$\sigma$ (sigma)	Serotonin-gated ion channels 5-HT3	GVIIIA
$\tau$ (tau)	Somatostatin receptor	CnVA
$\chi$ (chi)	Neuronal noradrenaline transporter	MrIA, CMrVIA
$\omega$ (omega)	Voltage-gated Ca channels (blocker)	GVIA

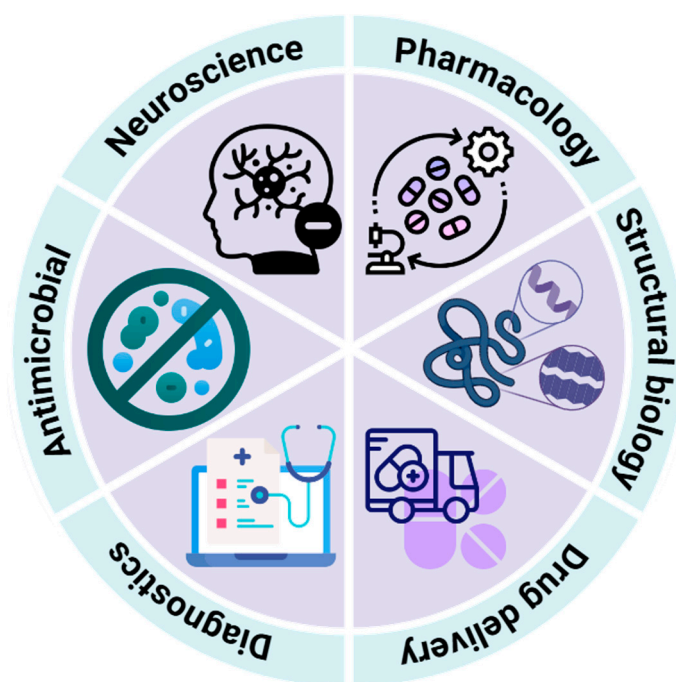
#### 2.4. Action and Applications of Conotoxins

Conotoxins, due to their structural stability, relatively small size, and excellent target specificity, serve as ideal molecular probes for target validation and peptide drug discovery [5]. Currently, conotoxins have become valuable research tools in neuroscience, pharmacology, biochemistry, structural biology, and molecular evolution (Figure 2). For example, in neuroscience, conotoxins are used to study neurotransmission [104,107]; in pharmacology, they are investigated as potential drugs, such as Prialt (ziconotide) for chronic pain [98,99]; in biochemistry, they aid in peptide biosynthesis studies [42,92,108]; and in molecular evolution, they are used to explore venom peptide evolution across *Conus* species [71,109]. They are also investigated as potential drugs, diagnostic agents, and drug leads [6]. By targeting specific ion channels, receptors, and other membrane proteins in the nervous system, conotoxins regulate neurotransmission and cellular functions, demonstrating broad potential for application in various fields.

Conotoxins hold significant potential in pain management, particularly in the treatment of neuropathic pain. Traditional pain management approaches, such as the use of opioid drugs, often lead to side effects such as addiction. In contrast, conotoxins, due to their high selectivity for specific ion channels, can effectively alleviate pain without the risk of addiction [4,6,99].  $\omega$ -conotoxin MVIIA (Ziconotide) is currently the only conotoxin-based drug approved by the FDA for refractory chronic pain. It exerts its effect by selectively blocking N-type calcium channels, thereby inhibiting neurotransmitter release and re-



ducing the transmission of pain signals [98]. The clinical application of Ziconotide has demonstrated the unique advantages of conotoxins in pain management, particularly in patients who do not respond to conventional treatments [99]. Recent studies have indicated that the  $\alpha$ -conotoxin RgIA derivative, RgIA4 (also known as KCP-400), acts as an antagonist of  $\alpha 9\alpha 10$  nAChRs, showing high efficacy in both humans and rodents. Moreover, its selective inhibition of  $\alpha 9\alpha 10$  nAChRs is at least 1000 times stronger than that of other drug targets, positioning it as a promising potential analgesic for the treatment of neuropathic pain [110–112]. The research on RgIA4 provides new directions for developing therapies for neuropathic pain, particularly for chronic pain induced by chemotherapy.



**Figure 2.** Key applications and potential of conotoxins. This figure highlights the diverse applications of conotoxins in various fields: (1) Pharmacology: Conotoxins are used in developing selective drugs for pain management, diabetes, and neurodegenerative diseases. (2) Neuroscience: By targeting specific ion channels and receptors, conotoxins regulate neurotransmission and cellular functions, advancing the study of neural signaling and circuits. (3) Antimicrobial:  $\alpha$ -conotoxin RgIA mutants demonstrate antimicrobial activity, offering new opportunities for antibiotic development. (4) Diagnostics: Conotoxins are explored as novel diagnostic tools for early disease detection. (5) Drug Delivery and Anticancer Therapies: The  $\alpha$ -ImI-paclitaxel conjugate shows promise in reducing tumor size and systemic toxicity, suggesting a new approach to cancer therapy. (6) Structural Biology: Conotoxins contribute to peptide biosynthesis research and receptor crystallization, supporting drug development and basic science.

Several studies have pointed to animal venom peptides as pharmacological tools and potential therapeutic agents for diabetes [53,55,113], with some peptides directly acting as mimics of endogenous metabolic hormones. Others target ion channels expressed in pancreatic  $\beta$ -cells [6]. Con-Ins G1, an insulin peptide derived from the venom of the geographic cone snail (*Conus geographus*) [114], can activate both fish and human insulin receptors. It is the true monomeric insulin, lacking the C-terminal B-chain required for dimerization [55,115], and as such, it can act more rapidly to reduce blood glucose levels. Another conotoxin, Konkunitzin-S1 (Conk-S1), derived from the venom of the striped cone snail (*Conus striatus*) [116], inhibits Kv1.7 delayed rectifier currents, leading to increased insulin secretion in rats following glucose stimulation of the pancreas. Unlike K<sub>ATP</sub> channel blockers, such as glibenclamide, Conk-S1 does not induce hypoglycemia, and

its effects on blood glucose are temporary, without altering basal glucose levels [55,117]. These characteristics make Conk-S1 a valuable tool for studying the role of Kv1.7 in insulin secretion and highlighting its potential as a treatment for type 2 diabetes (T2D).

Conotoxins, through their binding to ion channels on neuronal membranes, precisely regulate neuronal excitability, neurotransmitter release, and membrane potential changes. This ability has positioned them as powerful tools for investigating neuroelectrophysiology, signal transduction, and neural circuits [83,93]. It also can provide a theoretical foundation for the treatment of neurological disorders. These toxins have demonstrated significant potential in the treatment of neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease [93]. They can selectively modulate inter-neuronal communication and enhance synaptic function. Furthermore, the high specificity and adjustable structure of conotoxins make them ideal candidates for novel therapeutic agents for neurodegenerative conditions.

The potential of conotoxins in fields such as antimicrobial [118,119] and anti-cancer [99,120,121] therapies is gradually being explored, revealing broad application prospects. For instance, a study demonstrated that  $\alpha$ -ImI-conjugated paclitaxel, a chemotherapeutic drug, significantly reduced tumor mass in mice compared to the unconjugated control group while also decreasing systemic toxicity, highlighting the potential of conotoxins in drug delivery [120]. Additionally, research involving the synthesis of nine  $\alpha$ -conotoxin RgIA mutants using D-amino acids revealed their antimicrobial activity against pathogens and fungi by disrupting bacterial cell membranes. These mutants exhibited low hemolytic activity, reduced cytotoxicity, and enhanced stability, providing a novel approach to antibiotic development [119]. Furthermore,  $\alpha$ -conotoxins enhance the antitumor effects of cyclooxygenase and lipoxygenase inhibitors in Ehrlich carcinoma models in vitro and in vivo [122]. PnIA, RgIA, and ArIB11L16D potentiate the cytotoxic activity of baicalein and nordihydroguaiaretic acid, leading to reduced tumor growth and improved survival rates in mice [122]. These findings highlight the therapeutic potential of  $\alpha$ -conotoxins in cancer treatment.

As peptide molecules, conotoxins still face challenges commonly associated with peptide-based drugs. Currently, conotoxin-based pharmaceuticals on the market require intrathecal injection for direct delivery into the spinal cord, an invasive method. Therefore, the development of alternative, less invasive delivery methods, such as oral administration, represents a more ideal approach [94]. However, conotoxins possess several unique properties, including small molecular size, chemical diversity, and post-translational modifications, which offer considerable promise for both clinical and research applications. These features provide strong support for drug development and foundational scientific research. In addition, conotoxins have diverse roles across various fields, including medical diagnostics, structural biology, cosmetics, and agriculture, further highlighting their broad potential [6].

### 3. Conotoxin Databases and Resources

Conotoxin databases and resources play a crucial role in the field of bioinformatics, making significant contributions to sequence information retrieval, functional annotation, sequence analysis, evolutionary and phylogenetic studies, drug design and screening, as well as data sharing and collaboration. Traditional protein databases, such as UniProtKB, the Protein Data Bank, the Biological Magnetic Resonance Data Bank, and the protein databases provided by the National Center for Biotechnology Information (NCBI), are primary sources of molecular biology data. In previous studies, researchers have frequently utilized these resources to gather relevant information.

ConoServer (<https://www.conoserver.org/>, accessed on 26 December 2024), a specialized database for collecting conotoxin sequence and structural information, provides a wealth of data on conotoxins, making it a valuable resource for conotoxin research [28,29]. As of 1 December 2024, ConoServer has curated 3073 nucleotide sequences (from 90 *Conus* species), 8523 protein sequences (from 123 *Conus* species), and 247 3D structures (representing 14 cysteine frameworks from 48 *Conus* species). ConoServer organizes its data into a detailed classification scheme based on gene superfamilies, cysteine frameworks, and pharmacological families. These classifications, along with sequence annotations, provide an extensive set of features (such as amino acid composition, sequence motifs, and conserved regions), which are crucial for training machine learning models focused on peptide sequence analysis. The regularly updated data in ConoServer aid in the creation of high-quality datasets for classification and prediction and has been widely used to predict the classes and properties of conotoxins, including toxicity, ion channel targeting, and receptor binding.

The quantities of various classes in benchmark datasets based on different classification criteria, along with the associated references, are presented in Table 3.

**Table 3.** The benchmark datasets of conotoxin superfamily and ion channel-targeted conotoxin [79]. This table summarizes the number of conotoxin sequences classified into various superfamilies and associated with specific ion channels. The data include both superfamily classification (A, M, O, T) and the types of ion channels targeted by conotoxins (K, Na, Ca), along with the total number of sequences available in each category.

Superfamily						
	A	M	O	T	Total Number	Reference
S1	24	43	45	17	116	[16,19,123,124]
S2	26	49	70	55	216	[17,18]
Type of Ion Channel						
	K-Conotoxin	Na-Conotoxin	Ca-Conotoxin		Total Number	Reference
I1	24	43	45		112	[39,125–127]
I2	26	49	70		145	[128]

The first benchmark dataset for conotoxin superfamilies, called S1, includes 116 mature conotoxin sequences, along with a negative dataset containing 60 short peptide sequences that do not belong to any of the four superfamilies (A, M, O, or T) [123]. The second benchmark dataset, S2, expands upon this by including 261 entries from SwissProt that cover the four superfamilies, providing a broader and more diverse resource for classification tasks [17].

In addition to gene family-based datasets, functional datasets focusing on ion channel-targeting conotoxins have also been developed. The non-redundant benchmark dataset I1 includes 112 mature conotoxins, comprising 24 K-conotoxins, 43 Na-conotoxins, and 45 Ca-conotoxins [125]. The I2 dataset provides an extended set of ion channel-targeting conotoxins with 145 sequences [128]. These functional datasets are all curated from UniProt.

Many studies have constructed additional conotoxin datasets through rigorous filtering and management steps to align with the classification objectives of the models, including the classification of gene superfamilies [21,22], prediction of novel conotoxins [22–24] and prediction of conotoxin classes and conotoxins that target nAChRs [40], as illustrated in Table 4.

**Table 4.** Additional conotoxin datasets. This table lists conotoxin-related datasets, including their source, number of sequences, main purpose, and reference. These datasets are used for tasks like superfamily classification and prediction of new conotoxins.

Model	Source	Dataset	Main Purpose	Reference
ConoSorter	ConoServer, UniProt	2008 sequences: 1390 from the conotoxin superfamilies and 1931 from the classes.	Superfamily Classification.	[21]
ConoDictor 2.0	UniProt	727 sequences (from 19 superfamilies)	Superfamily Classification and Prediction of New Conotoxins	[22]
ConusPipe	ConoServer, UniProt	4950 conotoxin sequence; 52,613 non-conotoxin transcripts	Prediction of New Conotoxins	[23]
	ConoServer, PDB, BMRB	Small P: 154; Extended P: 184 Small EN: 180; Extended EN: 560 Small HN: 178; Extended HN: 317	Prediction of New Conotoxins	[24]
	PDB, BMRB	102 nAChR binders (98 alpha/29 mu/21 omega and others) 82 non-nAChR binders	Predict conotoxin classes and conotoxins that target nAChRs	[40]

Table note: P: positive dataset contains conotoxins; EN: easy-negative dataset contains non-toxic peptides; HN: hard-negative dataset contains toxic peptides.

## 4. Sequence-Based Classification Prediction and Generation Research

### 4.1. Classification Tools

In recent years, researchers have integrated transcriptomics, proteomics, biochemistry, physiology, and bioinformatics into the field of venomomics to accelerate the discovery of conotoxins [1]. Venomomics involves the prediction and classification of putative conotoxins at the transcriptomic level. Traditionally, this process has relied on time-consuming BLAST [76] searches of specialized databases, but homology-based searches are not always accurate in predicting the activity and function of toxins, often resulting in incomplete or false-positive results. As a complement to BLAST, tools such as ConoPrec [28] provide relevant hints for superfamily and family classification based on the signal peptide sequence of the submitted precursor. However, these tools still face issues when signal sequences are missing, making predictions impossible. With the expansion of machine learning applications in protein classification, several tools that combine machine learning methods have been developed to address these limitations.

#### 4.1.1. Conotoxin Superfamily Classification Models

In 2017, Dao et al. [79] systematically reviewed the progress of machine learning in the classification of conotoxins and provided a detailed description of the entire classification workflow. In the early stages of applying machine learning to conotoxin research, classification models were primarily focused on predicting conotoxin superfamily types. Most of these models were introduced in their review, though there has been relatively less research in this area in recent years. This paper highlights the updated version of ConoDictor 2.0, along with other classifiers that also use Profile Hidden Markov Models (pHMM) to capture sequence characteristics and compare their performance.

The Profile Hidden Markov Model (pHMM) offers a flexible probabilistic framework that can capture changes in sequences such as insertions and deletions [77]. By training on multiple sequence alignments and utilizing known sequence data to construct the model, pHMMs can identify both conserved and variable regions within sequences, making them well-suited for representing biological sequences with diverse and complex structures [129].

The effectiveness of this method in capturing sequence features of other proteins has been well-established [130–132]. Silja Laht et al. [15] developed a computational method based on pHMM to predict and classify all described conotoxin superfamilies and families to identify new conotoxins from *Conus* genomes or transcriptomes. They constructed independent models based on 24 described conotoxin categories (16 superfamilies and 8 families), focusing on signal, propeptide, and mature regions. This approach overcame the issue of signal peptide absence and ultimately led to the development of 62 pHMM models, most of which demonstrated high sensitivity and specificity even when trained on small sample sizes.

ConoSorter is a sequence pattern search and discovery pipeline based on regular expressions and pHMMs [21]. The process begins with a stringent screening using regular expressions to filter out sequences that match known patterns. Then, the pHMM is employed to validate and identify more complex sequences that were mismatched initially. This hierarchical filtering strategy allows ConoSorter to perform an initial classification of sequences and then refine the analysis of edge sequences using more complex models. This approach can improve computational efficiency and enhance the capability to handle large-scale datasets. Additionally, ConoSorter offers a significant advantage in discovering new toxins compared to previous tools. However, the program still requires substantial manual input from users for result classification, especially when dealing with novel or atypical sequences.

ConoDicator combines two complementary feature-capturing methods, the Position-Specific Scoring Matrix (PSSM) and Profile Hidden Markov Model (pHMM), to build models for classification and prediction [22,133]. PSSM captures evolutionary characteristics within protein sequences by quantifying both their conserved and variable regions, helping to reveal evolutionary relationships, conserved functional domains, and evolutionary variations within protein families [134]. This method based on multiple sequence alignments of homologous sequences can calculate the frequency of each amino acid or nucleotide at each position and compare it with the background frequency to generate a score matrix. The matrix used to assess the similarity between a new sequence and conserved features is particularly useful for identifying conserved regions and patterns within the sequence [135,136].

ConoDicator enhances prediction accuracy by merging the match lists of pHMM and PSSM, applying weights based on the frequency of each predicted feature. The final classification result corresponds to the superfamily with the highest frequency. PSSM rapidly identifies conserved sequence patterns, which aids in initial screening and classification. On the other hand, the probabilistic framework of pHMM is more adept at capturing changes such as insertions or deletions in complex sequences, allowing for more precise classification. Therefore, the integration of both methods significantly improves classification sensitivity and specificity, particularly in cases where sequence data are incomplete or exhibit high variability.

The recently published ConoDicator 2.0 is an optimized and updated version of the original tool, providing an independent command-line utility with enhanced prediction accuracy and faster analysis speed [22]. The original version of ConoDicator [133] was limited by being web-based, which included potential bandwidth constraints and reduced flexibility for handling large datasets. The upgraded ConoDicator 2.0 addresses these limitations by supporting an offline, independent command-line interface along with multicore processing capabilities, providing improved prediction accuracy and faster analysis speeds. This upgrade lifts the restrictions on handling large transcriptome or proteome datasets, making it sufficiently capable of meeting the demands of large-scale *Conus* transcriptome or proteome analyses.



In terms of model training and optimization [137,138], the authors created a new benchmark dataset to eliminate erroneous annotations. This dataset was derived from UniProt, excluding the Conoserver data used in previous model training and retaining only complete precursors with superfamily annotations. The final dataset contained 727 sequences. The number of models in ConoDictor 2.0 increased from 48 in the original version to 158, covering more conotoxin superfamilies and variants. These models encompass known superfamilies and introduce new classifications. Therefore, ConoDictor 2.0 is better suited to adapt to highly complex sequences, such as those belonging to newly discovered superfamilies of conotoxins. Additionally, this version utilizes an optimized set of PSSM and pHMM, enabling efficient processing of the entire venom duct transcriptome while reducing false positives. By improving model structure and training data, ConoDictor 2.0 provides greater accuracy in handling sequence variability and incompleteness. It links identified conotoxins to the correct gene superfamily with an efficiency of nearly 99%. The establishment of multiple specific models for a single family significantly enhances specificity.

#### 4.1.2. Ion Channel-Targeted Conotoxin Classification Models

The classification of conotoxins into superfamilies can only provide indirect insights into their potential functions and cannot predict their specific receptor targets. For instance, both  $\Delta$ -conotoxin-like Ac6.1 and  $\Omega$ -conotoxin-like Ai6.2 belong to the O1 superfamily but target different types of ion channels. The  $\Delta$ -conotoxin-like Ac6.1 interacts with voltage-gated sodium channels, while the  $\Omega$ -conotoxin-like Ai6.2 inhibits voltage-gated calcium channels [139]. As a result, numerous machine learning-based prediction tools have emerged in this field, with detailed information on methods provided in Table 5.

**Table 5.** Comparison of different methods for the type of ion channel-targeted prediction. This table compares the performance of different prediction methods for ion channel-targeted conotoxins, including K-conotoxins, Na-conotoxins, and Ca-conotoxins. The metrics used for evaluation are Sensitivity (Sn), Accuracy (AA), and Overall Accuracy (OA), as well as the corresponding datasets and references for each method.

Model	Feature Quantity	Feature <sup>a</sup>	Feature Selection <sup>b</sup>	Algorithm <sup>c</sup>	Publication Year	Reference
RBF network	70	DPC	Binomial Distribution	RBF network	2013	[125]
iCTX-Type	50	PseAAC of DP Patterns	F-score, IFS	SVM (RBF Kernel)	2014	[126]
ICTCPred	503	Hybrid Features: CTD, g-Gap DC, PP, SSI	SMOTE, Relief, IFS	RF	2016	[128]
Fscore-SVM	180	PseAAC (newly added attributes)	F-score	SVM (RBF Kernel)	2016	[127]
AVC-SVM	68	DPC	ANOVA, PCC	SVM	2017	[140]
ICTC-RAAC	156	NPC (AAC/DPC/TPC)	RAAC, ANOVA, IFS	SVM (RBF Kernel)	2020	[39]

Table note: <sup>a</sup>: AAC: amino acid composition; DPC: dipeptide composition; NPC: N-peptide combination; PseAAC: pseudo amino acid composition (newly added attributes:rigidity/flexibility/irreplaceability); CTD: Composition, Transition, Distribution features; G-gap DC: Gapped-Dihedral Correlation; PP: Physicochemical Properties; SSI: Secondary Structure Information. <sup>b</sup>: IFS: Incremental Feature Selection; SMOTE: Synthetic Minority Over-sampling Technique; ANOVA: Analysis of Variance; PCC: Pearson Correlation Coefficient; RAAC: Reduced Amino Acid Composition. <sup>c</sup>: RF: Random Forest; RBF: Radial Basis Function; SVM: Support Vector Machine.

Yuan et al. were the first to construct a non-redundant benchmark dataset I1 and predict the types of conotoxins targeting ion channels [125]. Using the dipeptide composition as a feature and optimizing the feature set based on the binomial distribution, they employed a radial basis function network to predict the types of conotoxins that target ion channels.

Subsequently, a series of studies were conducted building on this foundation, with most machine learning methods utilizing benchmark datasets for model training. Both iCTX-Type [126] and the model used pseudo-amino acid composition (PseAAC) [127] as a feature. They applied F-score-based algorithms for feature selection, with support vector machines (SVM) employed as the classification algorithm. However, three new amino acid residue properties, including rigidity, flexibility, and irreplaceability, were incorporated into the PseAAC. Their approach outperformed previous methods overall.

Two additional models based on this benchmark's dataset had progressed in feature selection and optimization. Analysis of Variance (ANOVA) was employed to classify ion channel-targeted conotoxins [140] to select the best features from the 400-dimensional dipeptide composition. It can calculate the ratio between-group to within-group variance for each attribute with a strong statistical foundation that allows for direct testing of feature differences. A study published in 2020 developed a model named ICTC-RAAC, which similarly used N-peptide combinations (including amino acid composition, dipeptide, and tripeptide combinations) to represent conotoxin sequences. This model employed ANOVA combined with Incremental Feature Selection (IFS) to enhance prediction performance and help select the discriminative features based on their F-values. It improved network predictors and introduced a Reduced Amino Acid Cluster (RAAC) approach to simplify protein complexity [141,142]. RAAC clustered the 20 naturally occurring amino acids into groups based on their similar physicochemical properties, resulting in a simplified alphabet [143–145]. By reducing protein complexity, this method helps identify conserved sequence regions within proteins (Table 6).

**Table 6.** Comparison of performance for different types of ion channel-targeted prediction.

Methods	Dataset	Sn			AA	OA	Reference
		K-Conotoxin	Na-Conotoxin	Ca-Conotoxin			
RBF network	I1	0.917	0.884	0.889	0.897	0.893	[125]
iCTX-Type	I1	0.833	0.978	0.898	0.903	0.911	[126]
ICTCPred	I2	1	0.919	1	0.973	0.957	[128]
Fscore-SVM	I1	0.917	0.953	0.953	0.942	0.946	[127]
AVC-SVM	I1	0.931	0.942	0.892	0.922	0.920	[140]
ICTC-RAAC	I1	0.917	0.954	10		0.964	[39]

The major models employ Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel as the classification algorithm. Support Vector Machines (SVM), initially developed by Vapnik et al. [146], are especially well-suited for small sample sizes, high-dimensional data, and complex pattern recognition tasks. SVM improves classification accuracy by maximizing the margin between classes and it can effectively avoid overfitting even with small datasets. It has been used in many practical applications, such as image recognition and text classification. The hierarchical use of SVM allows for the effective decomposition and resolution of complex classification tasks. Therefore, several studies have employed SVM to predict conotoxin types. The basic principle of SVM is to transform the input vectors into a high-dimensional Hilbert space and find a separating hyperplane in that space. The Gaussian Radial Basis Function (RBF) kernel [147,148] is a widely used kernel function due to its high performance in non-linear classification tasks.

A machine learning model, ICTCPred [128], was built based on I2 datasets for related predictions. The model uses hybrid features rather than a single discrete model to describe conotoxin samples. These hybrid features include 246 physicochemical properties of residues, obtained from APDbase, and incorporate composition, transition, and distribution (CTD), g-Gap dipeptide combinations (g-Gap DC), physicochemical properties (PP), and secondary structure information (SSI).

A notable challenge with this dataset is the imbalance among the different ion channel-targeting conotoxins, with Na<sup>+</sup>-channel-targeting conotoxins outnumbering Ca<sup>2+</sup>-channel-targeting conotoxins. To address this imbalance, ICTCPred employs the SMOTE technique (Synthetic Minority Over-sampling Technique) to balance the dataset, ensuring that the number of Na<sup>+</sup>-channel-targeting conotoxins is comparable to that of Ca<sup>2+</sup>-channel-targeting conotoxins.

Random Forest is a popular ensemble learning method applied to a wide range of biological prediction tasks [135,149–153]. It is capable of handling high-dimensional, complex, and noisy data and is effective in addressing issues with imbalanced datasets. By aggregating the results of multiple decision trees, Random Forest generates highly accurate classifiers, providing reliable predictions. In biology and medicine, it has proven to be an effective tool for disease prediction, gene analysis, and protein function prediction [154–158].

#### 4.1.3. Toxicity Prediction Models

It is estimated over 700 species of *Conus* worldwide [159], each capable of secreting more than 1000 distinct conotoxins [1]. The total number of different bioactive peptides produced by *Conus* is estimated to be around one million. However, to date, fewer than 0.1% of these conotoxins have been fully characterized [5,93]. Therefore, the discovery of novel conotoxins from the vast reservoir of these compounds using high-throughput and sensitive methods is crucial for the development of *Conus*-based pharmaceuticals.

Traditional methods for peptide toxicity prediction, including crude venom purification, cloning of conotoxin precursor genes, venom duct transcriptomics, and venom proteomics, often rely on experimental laboratory techniques. These methods are not only time-consuming and labor-intensive but also costly [3]. In contrast, machine learning provides an automated, efficient, and precise solution that makes the toxin discovery process faster, more accurate, and scalable.

In recent years, several tools have been developed through deep learning or machine learning (ML) techniques to predict the toxicity of peptides based on their primary amino acid sequences, thereby improving the toxicity identification process. The tools, such as ClanTox [160], ToxClassifier [23], ToxPred [161], and TOXIFY [162], are constructed to distinguish venom proteins from non-venom proteins, offering valuable insights into conotoxin recognition. Published conotoxin identification models, such as ConusPipe [23] and ConoDictor 2.0 [22], have significant improvements in the machine learning models and feature selection to some extent enhanced prediction accuracy.

ConusPipe utilizes the chemical features of conotoxins to predict toxin transcripts in the *Conus* transcriptome, overcoming the limitations of traditional homology-based search methods (such as BlastX [28,29] and HMMER [21,163]). Even conotoxins that evolve rapidly and exhibit significant sequence divergence retain three conserved regions: the N-terminal signal sequence, the pro-peptide region, and the mature toxin region [8,164,165]. Based on these characteristics, ConusPipe constructs machine learning models and extracts 16 features (including SignalP D-value, cysteine percentage, molecular weight, the percentage of positively and negatively charged amino acids, and isoelectric point) from 4950 known toxin sequences and 52,613 non-toxic transcripts. Given the limited training data and a large dataset, three machine learning models are employed: logistic regres-

sion [166], effective in binary classification; semi-supervised learning (label spreading), which leverages unlabeled true positives; and artificial neural networks (perceptron), which perform well on large training sets, and are combined with cross-species Blastp to prevent the omission of potential toxin sequences.

When tested on 540,261 protein sequences, the ConusPipe model combination achieved a specificity of 99.92% in distinguishing conotoxins from other proteins. After performing homologous sequence filtering with the Uniprot and ConoServer databases, the remaining transcripts were input into the model for prediction, identifying new candidate toxins. Ultimately, from the further filtered 5148 sequences, 187 were recognized by all four models, and 709 were recognized by three models, validating the effectiveness and advantage of the multi-model combination in toxin discovery.

The tools mentioned in the previous section, Conosorter [21,78] and ConoDictor 2.0 [22], are primarily designed to classify conotoxins from different superfamilies, but they can also identify whether a sequence is a conotoxin. ConoDictor was tested using a dataset containing known, correctly annotated conotoxins, along with sequences from random toxins (spiders, snakes, sea urchins, jellyfish) from UniProtKB. The results showed that ConoDictor 2.0 outperforms Conosorter v1.1 and BLAST v2.9.0+ in distinguishing conopeptides, with the highest specificity.

Conotoxin proteins typically feature flexible peptide backbones, a high proportion of cysteine-linked disulfide bond patterns, and rich post-translational modifications, characteristics that pose significant challenges for toxin prediction. Most current prediction methods and models neglect the impact of disulfide bond patterns and certain post-translational modifications on the prediction outcomes. In recent research, integrated physicochemical features (P), secondary structure features (SS), post-translational modification (P2) and collision cross-section (CCS) [24] significantly enhanced the model's performance in distinguishing between the toxin and non-toxic peptides. Among these, the collision cross-section (CCS) is a novel feature introduced for conotoxin prediction. CCS values, derived from ion mobility spectrometry (IM-MS) experiments or high-performance collision cross-section software (HPCCS) [167], represent experimental measurements that depend on factors such as molecular size, shape, charge, and polarization. The inclusion of this feature supplements the missing three-dimensional structural information and molecular interaction characteristics in existing conotoxin prediction methods. The CCS reflects the conformational differences between isomers of the same chemical composition formed by different disulfide bonds in conotoxins, thus aiding in the assessment of their varying toxicities.

Additionally, the authors classified negative samples into “easy negatives” (non-toxic samples) and “hard negatives” (toxic peptides from other species), and used a combination of machine learning models—penalized logistic regression (PLR), support vector machine (SVM), random forest (RF), and XGBoost—for training, which improved the model's ability to distinguish complex negative samples. By employing multi-level sample classification and integrating diverse features, Monroe et al. significantly enhanced the accuracy of their model. In particular, random forests demonstrated excellent performance across multiple datasets, with higher overall accuracy (OA) and recall rates than the best existing method in the field, ToxinPred [161], especially in distinguishing between toxin and non-toxin proteins. When all features (P + SS + CCS + P2) were combined, all four classifiers performed optimally, with a converged OA of approximately 96%, an AA of 95%, and an F1 score of 0.92. For “hard negatives”, the model exhibited high robustness and precision, making the identification of conotoxins more reliable. Monroe et al.'s research, through the introduction of novel features and improved classification strategies, successfully achieved precise conotoxin prediction in complex contexts, laying a foundation for further understanding their biological functions and potential in drug development.

Overall, from the integrated classification capabilities of ConusPipe and ConoDictor 2.0 to the novel features and advanced classification models introduced by Monroe et al., significant breakthroughs have been made in the accuracy and applicability of conotoxin identification and classification methods. These studies provide robust tools and a solid theoretical foundation for large-scale toxin screening and future drug development. Additionally, toxicological testing methods, such as haemolytic assays and in vivo toxicity models, serve as valuable experimental validations for these prediction models. These traditional assays help assess the accuracy of the models by confirming whether peptides predicted to be toxic exhibit cytotoxicity [168,169], cause disruption to red blood cell membranes [104,168,169], or induce systemic toxicity in animal models [169]. For models that predict conotoxins, validation is further achieved through techniques such as mass spectrometry (MS), transcriptomics, proteomics, and structural biology analyses [74,92,170]. These methods ensure that the predicted peptides are conotoxins by confirming their identity and structural features. By integrating these experimental approaches with computational models, the accuracy of toxicity predictions can be significantly enhanced [171], providing a more reliable foundation for the identification and development of therapeutic agents.

#### 4.2. Deep Learning-Based Generation and Prediction Framework

In recent years, deep learning (DL) has made significant advances in molecular generation and prediction. Tools such as ProteinGAN [172], ProGen [173], ProtGPT [174], and ProteinMPNN [175] have achieved tremendous success in generating proteins and peptides. Compared to traditional methods for discovering conotoxins, such as purifying toxins from venom or modifying toxin structures, deep learning approaches leverage large-scale data to learn and generate novel toxin molecules, greatly accelerating the discovery of new toxins. Against this backdrop, the recently proposed ConoDL framework [176] represents a breakthrough in this field. ConoDL consists of two core modules: the Conotoxin Generation Model (ConoGen) and the Conotoxin Prediction Model (ConoPred), which provide innovative tools and methods for the study and application of conotoxin.

ConoGen is based on the Transformer architecture and utilizes transfer learning techniques, employing the large language model ProGen to generate conotoxins. The Transformer model is a pivotal architecture in modern deep learning, widely applied across various fields such as language understanding, image processing, and information retrieval. Its core feature, the self-attention mechanism, connects all tokens in a sequence through a correlation-based pooling operation [177,178]. Due to its inherent advantages in processing sequential data, the Transformer model effectively captures the sequence characteristics of biological molecules like DNA, RNA, and proteins. While its attention mechanism helps identify long-term dependencies within complex molecular structures [178]. In ConoGen, the Transformer architecture is employed to learn the intricate internal patterns of conotoxins and generate artificial sequences resembling natural conotoxins. The successful application of this approach significantly enhances the generation and prediction capabilities of conotoxins.

ConoPred, based on the Wasserstein Autoencoder (WAE), was used for the prediction and screening of generated conotoxins, effectively removing sequences that are unlikely to be functional toxins and thus further improving the efficiency of the research process. The WAE model typically constrains the output of the encoder using a distribution, such as the standard normal distribution, ensuring that the representations in the latent space align. This constraint is achieved by minimizing the Wasserstein distance between the encoder's output distribution and the distribution, enabling the model to distinguish which sequences are close to the data distribution and which deviate from the normal range [179]. As a result,



the WAE model effectively captures the semantic relationships between peptide sequences, accelerating the discovery of functional peptides. It has been successfully applied to study antimicrobial peptides (AMPs) [180] and anticancer peptides (ACPs) [181]. The WAE plays a central role in feature extraction and generating probability scores. Its main function is to represent the input conotoxin sequences in the latent space and learn how to effectively encode and decode these sequences through reconstruction loss and prior constraints. Ultimately, the model uses this latent representation to assess whether the input sequence possesses features similar to real conotoxins, generating a probability score that indicates the likelihood of the sequence being a genuine conotoxin.

Experiments have demonstrated that the artificial conotoxins generated by ConoDL are like natural toxins in both sequence and spatial structure, and they exhibit novel cysteine scaffolds. These findings hold promise for advancing the molecular space exploration of conotoxins and the discovery of drug-active variants.

Emerging deep-learning techniques show significant potential in protein classification [182,183] and generation tasks [184,185]. Methods such as diffusion models [186], Generative Adversarial Networks (GANs) [187], Conditional Variational Autoencoders (CVAE) [188], multimodal Transformers [189], and Graph Neural Networks (GNNs) [190] have demonstrated advantages in generating diverse toxin sequences, enhancing prediction accuracy, and performing feature learning. The integration of generation and optimization frameworks with reinforcement learning is expected to lead to breakthroughs in conotoxin research. These methods not only improve the efficiency of new toxin discovery but also provide strong technical support and development potential for the application of toxins in drug development and neuroscience.

## 5. Future Research Directions

### 5.1. Optimization and Expansion of Data Resources

Current databases, such as ConoServer, have significantly contributed to conotoxin research by providing extensive sequence and structural information [28,29]. However, significant gaps remain in the coverage of rare superfamilies and novel variants of conotoxins [21,80]. Many of these rare or less-studied conotoxins have not been fully characterized or incorporated into existing databases, limiting the potential for comprehensive analysis and discovery. Additionally, the rapid evolution and high diversity of conotoxin sequences pose challenges for maintaining up-to-date and accurate databases [45,46].

The issue of constructing datasets is data imbalance. Certain conotoxins, such as those targeting sodium channels, tend to dominate the existing data, which may affect the accuracy of prediction models. Advanced sampling techniques have been employed to synthesize minority class samples, with some existing models already utilizing the SMOTE method to generate new samples [24,40,128]. Another approach involves using weighted loss functions to assign greater weight to minority class samples while integrating ensemble learning methods can effectively mitigate the problem of class imbalance [191,192].

Furthermore, the function of conotoxins is influenced not only by their amino acid sequences but also by factors such as their three-dimensional structures and post-translational modifications [193,194]. Future research could consider integrating sequence data with structural data (e.g., NMR, X-ray crystallography), mass spectrometry data, and other modal information to build more [28,29] robust hybrid models. For example, incorporating information about the three-dimensional structure of the toxins and their interactions with receptors into the model could significantly improve its ability to predict toxin function, particularly for functional prediction tasks, where structural data can provide crucial supplementary information.

### 5.2. Introduction of Multi-Modal Features

Current conotoxin classification and prediction models primarily rely on sequence information, such as amino acid composition and dipeptide patterns, with limited use of structural features and post-translational modifications [17,18,20,39,125,126,128,133,140]. Integrating multi-modal features, such as collision cross-section (CCS) and molecular surface charge distribution, could significantly enhance model performance. However, the development and application of such features in conotoxin research are still in their infancy. As our understanding of the mechanisms of action of conotoxins deepens, new features, such as collision cross-section (CCS), molecular surface charge distribution, and molecular similarity, are likely to play a key role in enhancing the performance of prediction models. For instance, the “collision cross-section” (CCS) [24,40] combined with mass spectrometry data successfully increased the distinction between toxins and non-toxins. In future research, further integrating these new features with multi-modal data (e.g., mass spectrometry, NMR) could significantly enhance the accuracy and generalizability of models.

### 5.3. Application of Deep Learning Methods in Conotoxin Bioinformatics Research

With the rapid development of deep learning technologies, particularly in protein generation and biomolecular research, deep learning has become a critical tool in bioinformatics [195,196]. In the realm of prediction, deep learning can address data scarcity through transfer learning and pre-trained models, significantly improving the accuracy of predictions [197–199] based on training on large datasets. While deep learning has shown great potential in protein generation and prediction, its application in conotoxin research is still limited. Current models often neglect the integration of structural biology and molecular dynamics simulations, which could provide crucial supplementary information for functional prediction. Future research should explore advanced deep learning techniques, such as multi-task learning and reinforcement learning, to improve conotoxin generation and classification. Deep learning models can automatically capture sequence characteristics and structural features of conotoxins, especially in complex toxins, allowing for the detection of subtle structural differences and the prediction of potential targets. Multi-task learning (MTL) has emerged to tackle the challenges of data diversity and structure complexity and enhance model adaptability and generalization. MTL has already shown promising results in various protein prediction models [200–202] and, when applied to conotoxins, can optimize models across multiple dimensions by jointly training tasks such as toxin type classification, function prediction, and toxicity assessment, thereby improving the ability to handle complex data.

In terms of generation, deep learning also exhibits immense potential. The algorithms, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can discover new toxin molecules and explore uncharted molecular spaces. For instance, ConoDL, which employs a Transformer architecture to generate conotoxins, demonstrates the vast potential of deep learning in molecular generation [176]. These techniques, integrating known structural and functional data, not only enable the discovery of new toxin molecules but also facilitate the design of molecules with specific functions, providing possible candidates for new drug development.

### 5.4. Development of Diverse Models

Current classification tools primarily focus on ion channel targets, neglecting other potential targets such as nicotinic acetylcholine receptors (nAChRs) and G-protein coupled receptors (GPCRs) [39,125–127,140]. Additionally, quantitative structure–activity relationship (QSAR) models, which could predict the bioactivity and toxicity of new toxins, are

underdeveloped. Future studies should expand to include multi-receptor target prediction and QSAR models to enhance the generalizability and accuracy of conotoxin research.

Although there has been progress in developing classification models for conotoxins targeting different ion channels, most research has focused on ion channel targets [39,125–127,140], neglecting other potential targets, such as nicotinic acetylcholine receptors (nAChRs), N-methyl-D-aspartate (NMDA) receptors, and G-protein coupled receptors (GPCRs). Conotoxins exert their effects by binding to these neurotransmitter receptors, modulating the transmission of nerve signals, and offering significant potential for their application in neuroscience and drug development. Some studies have begun to explore these non-ion channel receptors, such as predicting conotoxins that target acetylcholine receptors [24,40], but research in this area remains limited. Therefore, future studies should expand to include multi-receptor target prediction and classification models, using advanced algorithms such as multi-task learning and deep learning to identify the targeting of conotoxins to different receptors, thereby enhancing the generalizability and accuracy of the models. In addition, Raman spectroscopy combined with principal component analysis (PCA) has been explored for rapid and precise classification of conotoxins based on structural features, distinguishing between toxin types and even disulfide isomers [203]. With the accumulation of experimental data and the advancement of computational capabilities, integrating conotoxins structural biology, molecular dynamics simulations, and high-throughput screening technologies can further improve receptor-target classification models, providing a more accurate theoretical foundation for drug development and disease treatment applications.

Current classification tools primarily rely on extracting sequence information [204], often overlooking the spatial structure and functional characteristics. Given that conotoxins typically have complex three-dimensional structures and various post-translational modifications. Incorporating QSAR models will provide more precise information for toxin function prediction and drug design. QSAR is a predictive model based on the quantitative relationship between molecular structure and biological activity [205]. In QSAR models, the structural features of molecules (such as amino acid composition, disulfide bond patterns, molecular size, charge distribution, etc.) are statistically related to their biological activities (such as receptor affinity, inhibition, etc.), enabling the prediction of the activity of untested molecules [206,207]. For conotoxins, QSAR models can assist researchers in understanding how different structural features influence their interaction with specific targets. For example, some toxins act by binding to specific ion channels, while others may target acetylcholine receptors [42]. By constructing QSAR models, researchers can predict the potential targets and biological activities of unknown toxins, thus accelerating the discovery of novel drugs. Constructing high-quality QSAR models requires sufficient structural data and accurate biological activity data. With advances in structural biology techniques and the updating data resources, increasing numbers of conotoxin three-dimensional structural data are becoming available, providing the necessary structural information for QSAR modeling. Biological activity data typically include information on the toxin's affinity for specific receptors, toxicity (e.g., LD50 values), neuroactivity (e.g., impact on neurotransmission), and inhibition of special effect ion channels. Currently, biological activity data are relatively scarce, while the related data through ongoing research will support the development of more robust QSAR models.

## 6. Conclusions

Conotoxins are potent peptides from cone snail venom known for modulating ion channels, receptors, and neurotransmitter systems. These molecules hold promise for drug development, especially in treating pain and neurological disorders.

We provide a detailed introduction to conotoxin research in bioinformatics, covering several key aspects: (1) Sequence Characteristics and Classification: Conotoxins feature unique cysteine-rich frameworks and disulfide bond patterns, essential for their function and classification. They are categorized into superfamilies and pharmacological families based on sequence homology and functional properties, aiding in understanding their mechanisms and therapeutic potential. (2) Databases and Resources: Significant progress has been made in creating conotoxin databases like UniProt and ConoServer, which provide extensive information on sequences, structures, and functions. However, gaps remain, particularly for rare or less studied toxin variants. (3) Machine Learning and Deep Learning in Conotoxin Research: Bioinformatics tools using machine learning (ML) and deep learning (DL) have advanced conotoxin research. Tools predict superfamilies and ion channel targets, while deep learning frameworks like ConoDL enable conotoxin sequence generation and improve function, toxicity, and receptor predictions.

As conotoxin research continues to evolve, several key areas hold promise for future advancements. These trends include the expansion of conotoxin databases, the incorporation of multi-modal data (e.g., structural and mass spectrometry data) into prediction models, and the application of advanced deep learning techniques [208–210] to improve conotoxin generation, classification, and function prediction. Furthermore, developing more diverse models such as protein language models [186,187] and deep forest models [188] for receptor prediction and quantitative structure–activity relationships (QSAR) will contribute to targeted therapies and drug discovery.

**Author Contributions:** Conceptualization, resources, and data curation, R.L., H.L. and K.D.; methodology, formal analysis, investigation, and visualization, R.L., J.Y. and D.Y.; validation, S.L. and H.Z.; writing—original draft preparation, R.L. and J.F.; writing—review and editing, H.L., J.F. and K.D.; supervision, project administration, and funding acquisition, K.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (grant numbers 62372090, 82130112).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study.

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

## References

1. Dutertre, S.; Jin, A.H.; Kaas, Q.; Jones, A.; Alewood, P.F.; Lewis, R.J. Deep venomomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteom.* **2013**, *12*, 312–329. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3567856/> (accessed on 26 December 2024). [CrossRef] [PubMed]
2. Margiotta, F.; Micheli, L.; Ciampi, C.; Ghelardini, C.; McIntosh, J.M.; Di Cesare Mannelli, L. *Conus regius*-derived conotoxins: Novel therapeutic opportunities from a marine organism. *Mar. Drugs* **2022**, *20*, 773. Available online: <https://www.mdpi.com/1660-3397/20/12/773> (accessed on 26 December 2024). [CrossRef]
3. Fu, Y.; Li, C.; Dong, S.; Wu, Y.; Zhangsun, D.; Luo, S. Discovery methodology of novel conotoxins from *Conus* species. *Mar. Drugs* **2018**, *16*, 417. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6266589/> (accessed on 26 December 2024). [CrossRef] [PubMed]
4. Pennington, M.W.; Czerwinski, A.; Norton, R.S. Peptide therapeutics from venom: Current status and potential. *Bioorganic Med. Chem.* **2018**, *26*, 2738–2758. Available online: <https://www.sciencedirect.com/science/article/pii/S096808961731653X> (accessed on 26 December 2024). [CrossRef]
5. Jin, A.H.; Muttenthaler, M.; Dutertre, S.; Himaya, S.W.A.; Kaas, Q.; Craik, D.J.; Lewis, R.J.; Alewood, P.F. Conotoxins: Chemistry and biology. *Chem. Rev.* **2019**, *119*, 11510–11549. [CrossRef]

6. Bjørn-Yoshimoto, W.E.; Ramiro, I.B.L.; Yandell, M.; McIntosh, J.M.; Olivera, B.M.; Ellgaard, L.; Safavi-Hemami, H. Curses or cures: A review of the numerous benefits versus the biosecurity concerns of conotoxin research. *Biomedicines* **2020**, *8*, 235. [CrossRef] [PubMed]
7. Cruz, L.J.; Gray, W.R.; Olivera, B.M. Purification and properties of a myotoxin from *Conus geographus* venom. *Arch. Biochem. Biophys.* **1978**, *190*, 539–548. [CrossRef]
8. Bandyopadhyay, P.K.; Colledge, C.J.; Walker, C.S.; Zhou, L.M.; Hillyard, D.R.; Olivera, B.M. Conantokin-G precursor and its role in gamma-carboxylation by a vitamin K-dependent carboxylase from a *Conus snail*. *J. Biol. Chem.* **1998**, *273*, 5447–5450. [CrossRef]
9. Gray, W.R.; Luque, A.; Olivera, B.M.; Barrett, J.; Cruz, L.J. Peptide toxins from *Conus geographus* venom. *J. Biol. Chem.* **1981**, *256*, 4734–4740. Available online: <https://www.sciencedirect.com/science/article/pii/S0021925819693130> (accessed on 26 December 2024). [CrossRef]
10. Phuong, M.A.; Mahardika, G.N. Targeted sequencing of venom genes from cone snail genomes improves understanding of conotoxin molecular evolution. *Mol. Biol. Evol.* **2018**, *35*, 1210–1224. [CrossRef] [PubMed]
11. Charoenkwan, P.; Schaduagratt, N.; Manavalan, B.; Shoombuatong, W. M3S-ALG: Improved and robust prediction of allergenicity of chemical compounds by using a novel multi-step stacking strategy. *Future Gener. Comput. Syst.* **2025**, *162*, 107455. [CrossRef]
12. Charoenkwan, P.; Schaduagratt, N.; Pham, N.T.; Manavalan, B.; Shoombuatong, W. Pretoria: An effective computational approach for accurate and high-throughput identification of CD8+ t-cell epitopes of eukaryotic pathogens. *Int. J. Biol. Macromol.* **2023**, *238*, 124228. [CrossRef]
13. Shoombuatong, W.; Homdee, N.; Schaduagratt, N.; Chumnanpuen, P. Leveraging a meta-learning approach to advance the accuracy of Nav blocking peptides prediction. *Sci. Rep.* **2024**, *14*, 4463. [CrossRef]
14. Shoombuatong, W.; Meewan, I.; Mookdarsanit, L.; Schaduagratt, N. Stack-HDAC3i: A high-precision identification of HDAC3 inhibitors by exploiting a stacked ensemble-learning framework. *Methods* **2024**, *230*, 147–157. [CrossRef] [PubMed]
15. Laht, S.; Koua, D.; Kaplinski, L.; Lisacek, F.; Stöcklin, R.; Remm, M. Identification and classification of conopeptides using profile hidden markov models. *Biochim. Et Biophys. Acta BBA Proteins Proteom.* **2012**, *1824*, 488–492. Available online: <https://www.sciencedirect.com/science/article/pii/S1570963911003219> (accessed on 26 December 2024). [CrossRef]
16. Lin, H.; Li, Q.Z. Predicting conotoxin superfamily and family by using pseudo amino acid composition and modified mahalanobis discriminant. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 548–551. Available online: <https://www.sciencedirect.com/science/article/pii/S0006291X07000459> (accessed on 26 December 2024). [CrossRef] [PubMed]
17. Fan, Y.X.; Song, J.; Kong, X.; Shen, H.B. PredCSF: An Integrated Feature-Based Approach for Predicting Conotoxin Superfamily. *Protein Pept. Lett.* **2011**, *18*, 261–267. Available online: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=0929-8665&volume=18&issue=3&page=261> (accessed on 26 December 2024). [CrossRef]
18. Yin, J.B.; Fan, Y.X.; Shen, H.B. Conotoxin Superfamily Prediction Using Diffusion Maps Dimensionality Reduction and Subspace Classifier. *Curr. Protein Pept. Sci.* **2011**, *12*, 580–588. Available online: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2037&volume=12&issue=6&page=580> (accessed on 26 December 2024). [CrossRef]
19. Zaki, N.; Wolfsheimer, S.; Nuel, G.; Khuri, S. Conotoxin protein classification using free scores of words and support vector machines. *BMC Bioinform.* **2011**, *12*, 217. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3133552/> (accessed on 26 December 2024). [CrossRef] [PubMed]
20. Koua, D.; Laht, S.; Kaplinski, L.; Stöcklin, R.; Remm, M.; Favreau, P.; Lisacek, F. Position-specific scoring matrix and hidden markov model complement each other for the prediction of conopeptide superfamilies. *Biochim. Et Biophys. Acta BBA Proteins Proteom.* **2013**, *1834*, 717–724. Available online: <https://www.sciencedirect.com/science/article/pii/S1570963912002956> (accessed on 26 December 2024). [CrossRef] [PubMed]
21. Lavergne, V.; Dutertre, S.; Jin, A.-h.; Lewis, R.J.; Taft, R.J.; Alewood, P.F. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genom.* **2013**, *14*, 708. Available online: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-708> (accessed on 26 December 2024). [CrossRef] [PubMed]
22. Koua, D.; Ebou, A.; Dutertre, S. Improved prediction of conopeptide superfamilies with ConoDicator 2.0. *Bioinform. Adv.* **2021**, *1*, vbab011. [CrossRef]
23. Li, Q.; Watkins, M.; Robinson, S.D.; Safavi-Hemami, H.; Yandell, M. Discovery of Novel Conotoxin Candidates Using Machine Learning. *Toxins* **2018**, *10*, 503. [CrossRef] [PubMed]
24. Monroe, L.K.; Truong, D.P.; Miner, J.C.; Adikari, S.H.; Sasiene, Z.J.; Fenimore, P.W.; Alexandrov, B.; Williams, R.F.; Nguyen, H.B. Conotoxin Prediction: New Features to Increase Prediction Accuracy. *Toxins* **2023**, *15*, 641. Available online: <https://www.mdpi.com/2072-6651/15/11/641> (accessed on 26 December 2024). [CrossRef]
25. Wang, Y.; Zhai, Y.; Ding, Y.; Zou, Q. SBSM-Pro: Support bio-sequence machine for proteins. *Sci. China Inf. Sci.* **2024**, *67*, 212106. [CrossRef]
26. Liang, C.; Wang, D.; Zhang, H.; Zhang, S.; Guo, F. Robust Tensor Subspace Learning for Incomplete Multi-View Clustering. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6934–6948. [CrossRef]



27. Pang, C.; Qiao, J.; Zeng, X.; Zou, Q.; Wei, L. Deep generative models in de novo drug molecule generation. *J. Chem. Inf. Model.* **2023**, *64*, 2174–2194. [CrossRef]
28. Kaas, Q.; Yu, R.; Jin, A.H.; Dutertre, S.; Craik, D.J. ConoServer: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **2011**, *40*, D325. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3245185/> (accessed on 26 December 2024). [CrossRef]
29. Kaas, Q.; Westermann, J.C.; Halai, R.; Wang, C.K.L.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2008**, *24*, 445–446. Available online: <https://academic.oup.com/bioinformatics/article/24/3/445/253535> (accessed on 26 December 2024). [CrossRef] [PubMed]
30. Kaas, Q.; Westermann, J.-C.; Craik, D.J. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon* **2010**, *55*, 1491–1509. [CrossRef]
31. Liu, M.; Li, C.; Chen, R.; Cao, D.; Zeng, X. Geometric Deep Learning for Drug Discovery. *Expert Syst. Appl.* **2023**, *240*, 122498. [CrossRef]
32. Liu, T.; Huang, J.; Luo, D.; Ren, L.; Ning, L.; Huang, J.; Lin, H.; Zhang, Y. Cm-siRPred: Predicting chemically modified siRNA efficiency based on multi-view learning strategy. *Int. J. Biol. Macromol.* **2024**, *264 Pt 2*, 130638. [CrossRef] [PubMed]
33. Wang, R.; Jiang, Y.; Jin, J.; Yin, C.; Yu, H.; Wang, F.; Feng, J.; Su, R.; Nakai, K.; Zou, Q. DeepBIO: An automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation and visualization analysis. *Nucleic Acids Res.* **2023**, *51*, 3017–3029. [CrossRef] [PubMed]
34. Tan, J.X.; Li, S.H.; Zhang, Z.M.; Chen, C.X.; Chen, W.; Tang, H.; Lin, H. Identification of hormone binding proteins based on machine learning methods. *Math. Biosci. Eng.* **2019**, *16*, 2466–2480. [CrossRef] [PubMed]
35. Zhang, H.Q.; Liu, S.H.; Li, R.; Yu, J.W.; Ye, D.X.; Yuan, S.S.; Lin, H.; Huang, C.-B.; Tang, H. MIBPred: Ensemble Learning-Based Metal Ion-Binding Protein Classifier. *ACS Omega* **2024**, *9*, 8439–8447. [CrossRef]
36. Wei, L.; He, W.; Malik, A.; Su, R.; Cui, L.; Manavalan, B. Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief. Bioinform.* **2020**, *22*, bbaa275. [CrossRef] [PubMed]
37. Xu, Y.; Liu, T.; Yang, Y.; Kang, J.; Ren, L.; Ding, H.; Zhang, Y. ACVPred: Enhanced prediction of anti-coronavirus peptides by transfer learning combined with data augmentation. *Future Gener. Comput. Syst.* **2024**, *160*, 305–315. [CrossRef]
38. Liu, T.; Qiao, H.; Wang, Z.; Yang, X.; Pan, X.; Yang, Y.; Ye, X.; Sakurai, T.; Lin, H.; Zhang, Y. CodLncScape Provides a Self-Enriching Framework for the Systematic Collection and Exploration of Coding LncRNAs. *Adv. Sci.* **2024**, *11*, 2400009. [CrossRef]
39. Sun, Z. ICTC-RAAC: An improved web predictor for identifying the types of ion channel-targeted conotoxins by using reduced amino acid cluster descriptors. *Comput. Biol. Chem.* **2020**, *89*, 107371. [CrossRef] [PubMed]
40. Truong, D.P.; Monroe, L.K.; Williams, R.F.; Nguyen, H.B. Machine learning framework for conotoxin class and molecular target prediction. *Toxins* **2024**, *16*, 475. Available online: <https://www.mdpi.com/2072-6651/16/11/475> (accessed on 26 December 2024). [CrossRef]
41. Woodward, S.R.; Cruz, L.J.; Olivera, B.M.; Hillyard, D.R. Constant and hypervariable regions in conotoxin propeptides. *EMBO J.* **1990**, *9*, 1015–1020. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC551772/> (accessed on 26 December 2024). [CrossRef]
42. Akondi, K.B.; Muttenthaler, M.; Dutertre, S.; Kaas, Q.; Craik, D.J.; Lewis, R.J.; Alewood, P.F. Discovery, synthesis and development of structure-activity relationships of conotoxins. *Chem. Rev.* **2014**, *114*, 5815–5847. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7610532/> (accessed on 26 December 2024). [CrossRef]
43. Robinson, S.D.; Safavi-Hemami, H.; McIntosh, L.D.; Purcell, A.W.; Norton, R.S.; Papenfuss, A.T. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS ONE* **2014**, *9*, e87648. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3914837/> (accessed on 26 December 2024). [CrossRef] [PubMed]
44. Jin, A.H.; Dekan, Z.; Smout, M.J.; Wilson, D.; Dutertre, S.; Vetter, I.; Lewis, R.J.; Loukas, A.; Daly, N.L.; Alewood, P.F. Conotoxin  $\Phi$ -MiXXVIIA from the superfamily G2 employs a novel cysteine framework that mimics granulin and displays anti-apoptotic activity. *Angew. Chem. Int. Ed.* **2017**, *56*, 14973–14976. Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201708927> (accessed on 26 December 2024). [CrossRef]
45. Chang, D.; Duda, T.F., Jr. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol. Biol. Evol.* **2012**, *29*, 2019–2029. [CrossRef]
46. Conticello, S.G.; Gilad, Y.; Avidan, N.; Ben-Asher, E.; Levy, Z.; Fainzilber, M. Mechanisms for evolving hypervariability: The case of conopeptides. *Mol. Biol. Evol.* **2001**, *18*, 120–131. [CrossRef]
47. Kumar, P.S.; Kumar, D.S.; Umamaheswari, S. A perspective on toxicology of *Conus* venom peptides. *Asian Pac. J. Trop. Med.* **2015**, *8*, 337–351. [CrossRef]
48. Zhang, Y.; Pan, X.; Shi, T.; Gu, Z.; Yang, Z.; Liu, M.; Xu, Y.; Yang, Y.; Ren, L.; Song, X.; et al. P450Rdb: A manually curated database of reactions catalyzed by cytochrome P450 enzymes. *J. Adv. Res.* **2024**, *63*, 35–42. [CrossRef] [PubMed]

49. Franco, A.; Pisarewicz, K.; Moller, C.; Mora, D.; Fields, G.B.; Marí, F. Hyperhydroxylation: A new strategy for neuronal targeting by venomous marine molluscs. In *Molluscs: From Chemo-Ecological Study to Biotechnological Application*; Cimino, G., Gavagnin, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 83–103, ISBN 978-3-540-30880-5. [CrossRef]
50. Ho, T.N.T.; Lee, H.S.; Swaminathan, S.; Goodwin, L.; Rai, N.; Ushay, B.; Lewis, R.J.; Rosengren, K.J.; Conibear, A.C. Posttranslational modifications of  $\alpha$ -conotoxins: Sulfotyrosine and C-terminal amidation stabilise structures and increase acetylcholine receptor binding. *RSC Med. Chem.* **2021**, *12*, 1574–1584. Available online: <https://pubs.rsc.org/en/content/articlelanding/2021/md/d1md00182e> (accessed on 26 December 2024). [CrossRef] [PubMed]
51. Wang, L.; Liu, J.; Ren, Z.; Chen, Y.; Xu, A. Discovery of two P-superfamily conotoxins, lt9a and lt9b, with different modifications on voltage-sensitive sodium channels. *Toxicon* **2017**, *134*, 6–13. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010117301587> (accessed on 26 December 2024). [CrossRef] [PubMed]
52. Ao, C.; Jiao, S.; Wang, Y.; Yu, L.; Zou, Q. Biological Sequence Classification: A Review on Data and General Methods. *Research* **2022**, *2022*, 0011. [CrossRef]
53. Lewis, R.J.; Dutertre, S.; Vetter, I.; Christie, M.J. *Conus* venom peptide pharmacology. *Pharmacol. Rev.* **2012**, *64*, 259–298. Available online: <https://pharmrev.aspetjournals.org/content/64/2/259> (accessed on 26 December 2024). [CrossRef]
54. Mir, R.; Karim, S.; Kamal, M.A.; Wilson, C.M.; Mirza, Z. Conotoxins: Structure, therapeutic potential and pharmacological applications. *Curr. Pharm. Des.* **2016**, *22*, 582–589. Available online: <https://www.eurekaselect.com/article/72077> (accessed on 26 December 2024). [CrossRef] [PubMed]
55. Robinson, S.D.; Safavi-Hemami, H. Venom peptides as pharmacological tools and therapeutics for diabetes. *Neuropharmacology* **2017**, *127*, 79–86. Available online: <https://www.sciencedirect.com/science/article/pii/S0028390817303088> (accessed on 26 December 2024). [CrossRef] [PubMed]
56. Liao, Y.; Peng, C.; Zhu, Y.; Fu, J.; Ruan, Z.; Shi, Q.; Gao, B. High conopeptide diversity in *Conus striatus*: Revealed by integration of two transcriptome sequencing platforms. *Front. Mar. Sci.* **2022**, *9*, 1060432. Available online: <https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2022.1060432/full> (accessed on 26 December 2024). [CrossRef]
57. Peng, C.; Yao, G.; Gao, B.-M.; Fan, C.-X.; Bian, C.; Wang, J.; Cao, Y.; Wen, B.; Zhu, Y.; Ruan, Z.; et al. High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing. *GigaScience* **2016**, *5*, 17. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832519/> (accessed on 26 December 2024). [CrossRef]
58. Li, X.; Liu, H.; Gao, C.; Li, Y.; Jia, D.; Yang, Y.; Yang, J.; Wei, Z.; Jiang, T.; Yu, R. ConoMode, a database for conopeptide binding modes. *Database: J. Biol. Databases Curation* **2020**, *2020*, baaa058. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7402919/> (accessed on 26 December 2024). [CrossRef] [PubMed]
59. Younis, S.; Rashid, S. Alpha conotoxin-BuIA globular isomer is a competitive antagonist for oleoyl-L-alpha-lysophosphatidic acid binding to LPAR6; a molecular dynamics study. *PLoS ONE* **2017**, *12*, e0189154. Available online: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189154> (accessed on 26 December 2024). [CrossRef]
60. Cordeiro, S.; Finol-Urdaneta, R.K.; Köpfer, D.; Markushina, A.; Song, J.; French, R.J.; Kopec, W.; de Groot, B.L.; Giacobassi, M.J.; Leavitt, L.S.; et al. Conotoxin  $\kappa$ M-R111, a tool targeting asymmetric heteromeric Kv1 channels. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1059–1064. Available online: <https://www.pnas.org/doi/abs/10.1073/pnas.1813161116> (accessed on 26 December 2024). [CrossRef]
61. Kohn, A.J.; Saunders, P.R.; Wiener, S. Preliminary studies on the venom of the marine snail *Conus*. *Ann. N. Y. Acad. Sci.* **1960**, *90*, 706–725. Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1960.tb26416.x> (accessed on 26 December 2024). [CrossRef]
62. Jin, H.; Cui, D.; Fan, Y.; Li, G.; Zhong, Z.; Wang, Y. Recent advances in bioaffinity strategies for preclinical and clinical drug discovery: Screening natural products, small molecules and antibodies. *Drug Discov. Today* **2024**, *29*, 103885. Available online: <https://www.sciencedirect.com/science/article/pii/S1359644624000102> (accessed on 26 December 2024). [CrossRef] [PubMed]
63. Chen, S.; Shen, C.; Li, W.; Fan, Y.; Yang, D.-H.; Wang, Y.; Feng, R.; Li, G.; Zhong, Z. Recent advances in bioactivity-guided drug screening strategies for pre-clinical and clinical drug discovery. *TrAC Trends Anal. Chem.* **2024**, *181*, 118042. Available online: <https://www.sciencedirect.com/science/article/pii/S0165993624005259> (accessed on 26 December 2024). [CrossRef]
64. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genom.* **2011**, *12*, 60. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3040727/> (accessed on 26 December 2024). [CrossRef] [PubMed]
65. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genom.* **2012**, *13*, 284. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3441800/> (accessed on 26 December 2024). [CrossRef] [PubMed]
66. Terrat, Y.; Biass, D.; Dutertre, S.; Favreau, P.; Remm, M.; Stöcklin, R.; Piquemal, D.; Ducancel, F. High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon* **2012**, *59*, 34–46. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010111003151> (accessed on 26 December 2024). [CrossRef]

67. Hackney, C.M.; Salcedo, P.F.; Mueller, E.; Koch, T.L.; Kjølgaard, L.D.; Watkins, M.; Zachariassen, L.G.; Tuelung, P.S.; McArthur, J.R.; Adams, D.J.; et al. A previously unrecognized superfamily of macro-conotoxins includes an inhibitor of the sensory neuron calcium channel Cav2.3. *PLOS Biol.* **2023**, *21*, e3002217. Available online: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002217> (accessed on 26 December 2024). [CrossRef]
68. Nguyen, L.T.T.; Craik, D.J.; Kaas, Q. Bibliometric review of the literature on cone snail peptide toxins from 2000 to 2022. *Mar. Drugs* **2023**, *21*, 154. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10058278/> (accessed on 26 December 2024). [CrossRef] [PubMed]
69. Espiritu, M.J.; Taylor, J.K.; Sugai, C.K.; Thapa, P.; Loening, N.M.; Gusman, E.; Baoanan, Z.G.; Baumann, M.H.; Bingham, J.-P. Characterization of the native disulfide isomers of the novel  $\chi$ -conotoxin PnID: Implications for further increasing conotoxin diversity. *Mar. Drugs* **2023**, *21*, 61. Available online: <https://www.mdpi.com/1660-3397/21/2/61> (accessed on 26 December 2024). [CrossRef] [PubMed]
70. Ebou, A.; Koua, D.; Addablah, A.; Kakou-Ngazoa, S.; Dutertre, S. Combined proteotranscriptomic-based strategy to discover novel antimicrobial peptides from cone snails. *Biomedicines* **2021**, *9*, 344. [CrossRef] [PubMed]
71. Lu, A.; Watkins, M.; Li, Q.; Robinson, S.D.; Concepcion, G.P.; Yandell, M.; Weng, Z.; Olivera, B.M.; Safavi-Hemami, H.; Fedosov, A.E. Transcriptomic profiling reveals extraordinary diversity of venom peptides in unexplored predatory gastropods of the genus *clavus*. *Genome Biol. Evol.* **2020**, *12*, 684–700. Available online: <https://academic.oup.com/gbe/article/12/5/684/5825281> (accessed on 26 December 2024). [CrossRef]
72. Hu, S.H.; Gehrmann, J.; Alewood, P.F.; Craik, D.J.; Martin, J.L. Crystal structure at 1.1 Å resolution of  $\alpha$ -conotoxin PnIB: Comparison with  $\alpha$ -conotoxins PnIA and GI. *Biochemistry* **1997**, *36*, 11323–11330. [CrossRef]
73. Hill, J.M.; Alewood, P.F.; Craik, D.J. Conotoxin TVIIA, a novel peptide from the venom of *Conus tulipa*. *Eur. J. Biochem.* **2000**, *267*, 4649–4657. Available online: <https://febs.onlinelibrary.wiley.com/doi/10.1046/j.1432-1327.2000.01507.x> (accessed on 26 December 2024). [CrossRef] [PubMed]
74. Wilson, D.T.; Bansal, P.S.; Carter, D.A.; Vetter, I.; Nicke, A.; Dutertre, S.; Daly, N.L. Characterisation of a novel  $\alpha$ -superfamily conotoxin. *Biomedicines* **2020**, *8*, 128. Available online: <https://www.mdpi.com/2227-9059/8/5/128> (accessed on 26 December 2024). [CrossRef] [PubMed]
75. Kang, T.S.; Jois, S.D.S.; Kini, R.M. Solution structures of two structural isoforms of CMrVIA  $\chi/\lambda$ -conotoxin. *Biomacromolecules* **2006**, *7*, 2337–2346. Available online: <https://pubs.acs.org/doi/10.1021/bm060269w> (accessed on 26 December 2024). [CrossRef]
76. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. Available online: <https://www.sciencedirect.com/science/article/pii/S0022283605803602> (accessed on 26 December 2024). [CrossRef]
77. Eddy, S.R. Profile hidden markov models. *Bioinformatics* **1998**, *14*, 755–763. Available online: <https://academic.oup.com/bioinformatics/article/14/9/755/259550> (accessed on 26 December 2024). [CrossRef]
78. Zamora-Bustillos, R.; Martínez-Núñez, M.A.; Aguilar, M.B.; Collí-Dula, R.C.; Brito-Domínguez, D.A. Identification of novel conotoxin precursors from the cone snail *Conus spurius* by high-throughput RNA sequencing. *Mar. Drugs* **2021**, *19*, 547. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8541002/> (accessed on 26 December 2024). [CrossRef]
79. Dao, F.Y.; Yang, H.; Su, Z.D.; Yang, W.; Wu, Y.; Ding, H.; Chen, W.; Tang, H.; Lin, H. Recent Advances in Conotoxin Classification by Using Machine Learning Methods. *Molecules* **2017**, *22*, 1057. [CrossRef]
80. Robinson, S.D.; Norton, R.S. Conotoxin gene superfamilies. *Mar. Drugs* **2014**, *12*, 6058–6101. Available online: <https://www.mdpi.com/1660-3397/12/12/6058> (accessed on 26 December 2024). [CrossRef] [PubMed]
81. Lee, S.F.; Davey, L. Disulfide bonds: A key modification in bacterial extracytoplasmic proteins. *J. Dent. Res.* **2017**, *96*, 1465–1473. Available online: <https://journals.sagepub.com/doi/10.1177/0022034517725059> (accessed on 26 December 2024). [CrossRef] [PubMed]
82. Fu, J.; Gao, J.; Liang, Z.; Yang, D. PDI-regulated disulfide bond formation in protein folding and biomolecular assembly. *Molecules* **2021**, *26*, 171. Available online: <https://www.mdpi.com/1420-3049/26/1/171> (accessed on 26 December 2024). [CrossRef] [PubMed]
83. Turner, A.; Kaas, Q.; Craik, D.J. Hormone-like conopeptides—New tools for pharmaceutical design. *RSC Med. Chem.* **2020**, *11*, 1235. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8126879/> (accessed on 26 December 2024). [CrossRef] [PubMed]
84. Biggs, J.S.; Watkins, M.; Puillandre, N.; Ownby, J.-P.; Lopez-Vera, E.; Christensen, S.; Moreno, K.J.; Bernaldez, J.; Licea-Navarro, A.; Corneli, P.S.; et al. Evolution of *Conus* peptide toxins: Analysis of *Conus californicus* reeve, 1844. *Mol. Phylogenetics Evol.* **2010**, *56*, 1–12. Available online: <https://www.sciencedirect.com/science/article/pii/S1055790310001429> (accessed on 26 December 2024). [CrossRef]
85. Elliger, C.A.; Richmond, T.A.; Lebaric, Z.N.; Pierce, N.T.; Sweedler, J.V.; Gilly, W.F. Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. *Toxicon* **2011**, *57*, 311–322. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010110004253> (accessed on 26 December 2024). [CrossRef] [PubMed]



86. Balsara, R.; Li, N.; Weber-Adrian, D.; Huang, L.; Castellino, F.J. Opposing action of conantokin-G on synaptically and extrasynaptically-activated NMDA receptors. *Neuropharmacology* **2012**, *62*, 2227–2238. Available online: <https://www.sciencedirect.com/science/article/pii/S002839081200038X> (accessed on 26 December 2024). [CrossRef]
87. Malmberg, A.B.; Gilbert, H.; McCabe, T.R.; Basbaum, A.I. Powerful antinociceptive effects of the cone snail venom-derived subtype-selective NMDA receptor antagonists conantokins G and T. *Pain* **2003**, *101*, 109. Available online: [https://journals.lww.com/pain/abstract/2003/01000/powerful\\_antinociceptive\\_effects\\_of\\_the\\_cone\\_snail.12.aspx](https://journals.lww.com/pain/abstract/2003/01000/powerful_antinociceptive_effects_of_the_cone_snail.12.aspx) (accessed on 26 December 2024). [CrossRef] [PubMed]
88. Hama, A.; Sagen, J. Antinociceptive effects of the marine snail peptides conantokin-G and conotoxin MVIIA alone and in combination in rat models of pain. *Neuropharmacology* **2009**, *56*, 556–563. Available online: <https://www.sciencedirect.com/science/article/pii/S0028390808004991> (accessed on 26 December 2024). [CrossRef]
89. Balsara, R.; Dang, A.; Donahue, D.L.; Snow, T.; Castellino, F.J. Conantokin-G attenuates detrimental effects of NMDAR hyperactivity in an ischemic rat model of stroke. *PLoS ONE* **2015**, *10*, e0122840. Available online: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122840> (accessed on 26 December 2024). [CrossRef]
90. Jacob, R.B.; McDougal, O.M. The M-superfamily of conotoxins: A review. *Cell. Mol. Life Sci.* **2010**, *67*, 17–27. [CrossRef] [PubMed]
91. Zhou, M.; Wang, L.; Wu, Y.; Zhu, X.; Feng, Y.; Chen, Z.; Li, Y.; Sun, D.; Ren, Z.; Xu, A. Characterizing the evolution and functions of the M-superfamily conotoxins. *Toxicon* **2013**, *76*, 150–159. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010113003863> (accessed on 26 December 2024). [CrossRef] [PubMed]
92. Ho, T.N.; Tran, T.H.; Le, H.S.; Lewis, R.J. Advances in the synthesis and engineering of conotoxins. *Eur. J. Med. Chem.* **2025**, *282*, 117038. Available online: <https://linkinghub.elsevier.com/retrieve/pii/S0223523424009206> (accessed on 26 December 2024). [CrossRef] [PubMed]
93. Gao, B.; Peng, C.; Yang, J.; Yi, Y.; Zhang, J.; Shi, Q. Cone snails: A big store of conotoxins for novel drug discovery. *Toxins* **2017**, *9*, 397. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5744117/> (accessed on 26 December 2024). [CrossRef]
94. Halai, R.; Craik, D.J. Conotoxins: Natural product drug leads. *Nat. Prod. Rep.* **2009**, *26*, 526–536. Available online: <https://pubs.rsc.org/en/content/articlelanding/2009/np/b819311h> (accessed on 26 December 2024). [CrossRef]
95. Green, B.R.; Bulaj, G.; Norton, R.S. Structure and function of  $\mu$ -conotoxins, peptide-based sodium channel blockers with analgesic activity. *Future Med. Chem.* **2014**, *6*, 1677–1698. [CrossRef] [PubMed]
96. Ekberg, J.; Craik, D.J.; Adams, D.J. Conotoxin modulation of voltage-gated sodium channels. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 2363–2368. Available online: <https://www.sciencedirect.com/science/article/pii/S1357272507003007> (accessed on 26 December 2024). [CrossRef] [PubMed]
97. Ramírez, D.; Gonzalez, W.; Fissore, R.A.; Carvacho, I. Conotoxins as tools to understand the physiological function of voltage-gated calcium (CaV) channels. *Mar. Drugs* **2017**, *15*, 313. Available online: <https://www.mdpi.com/1660-3397/15/10/313> (accessed on 26 December 2024). [CrossRef] [PubMed]
98. Pope, J.E.; Deer, T.R. Ziconotide: A clinical update and pharmacologic review. *Expert Opin. Pharmacother.* **2013**, *14*, 957–966. Available online: <http://www.tandfonline.com/doi/full/10.1517/14656566.2013.784269> (accessed on 26 December 2024). [CrossRef] [PubMed]
99. Safavi-Hemami, H.; Brogan, S.E.; Olivera, B.M. Pain therapeutics from cone snail venoms: From ziconotide to novel non-opioid pathways. *J. Proteom.* **2019**, *190*, 12–20. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6214764/> (accessed on 26 December 2024). [CrossRef] [PubMed]
100. Miljanich, G.P. Ziconotide: Neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.* **2004**, *11*, 3029–3040. Available online: <https://www.eurekaselect.com/62867/article> (accessed on 26 December 2024). [CrossRef] [PubMed]
101. Ferber, M.; Sporning, A.; Jeserich, G.; DeLaCruz, R.; Watkins, M.; Olivera, B.M.; Terlau, H. A novel *Conus* peptide ligand for K<sup>+</sup> channels\*. *J. Biol. Chem.* **2003**, *278*, 2177–2183. Available online: <https://www.sciencedirect.com/science/article/pii/S0021925819309688> (accessed on 26 December 2024). [CrossRef]
102. Ferber, M.; Al-Sabi, A.; Stocker, M.; Olivera, B.M.; Terlau, H. Identification of a mammalian target of  $\kappa$ M-conotoxin RIIIK. *Toxicon* **2004**, *43*, 915–921. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010104001370> (accessed on 26 December 2024). [CrossRef] [PubMed]
103. Hone, A.J.; McIntosh, J.M. Nicotinic acetylcholine receptors: Therapeutic targets for novel ligands to treat pain and inflammation. *Pharmacol. Res.* **2023**, *190*, 106715. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10691827/> (accessed on 26 December 2024). [CrossRef] [PubMed]
104. Sadeghi, M.; McArthur, J.R.; Finol-Urdaneta, R.K.; Adams, D.J. Analgesic conopeptides targeting G protein-coupled receptors reduce excitability of sensory neurons. *Neuropharmacology* **2017**, *127*, 116–123. Available online: <https://www.sciencedirect.com/science/article/pii/S0028390817302320> (accessed on 26 December 2024). [CrossRef] [PubMed]
105. Craig, A.G.; Norberg, T.; Griffin, D.; Hoeger, C.; Akhtar, M.; Schmidt, K.; Low, W.; Dykert, J.; Richelson, E.; Navarro, V.; et al. Contulakin-G, an O-glycosylated invertebrate neurotensin\*. *J. Biol. Chem.* **1999**, *274*, 13752–13759. Available online: <https://www.sciencedirect.com/science/article/pii/S002192581973230X> (accessed on 26 December 2024). [CrossRef]

106. Pei, S.; Wang, N.; Mei, Z.; Zhangsun, D.; Craik, D.J.; McIntosh, J.M.; Zhu, X.; Luo, S. Conotoxins targeting voltage-gated sodium ion channels. *Pharmacol. Rev.* **2024**, *76*, 828–845. [CrossRef] [PubMed]
107. Olivera, B.M.; Cruz, L.J.; De Santos, V.; LeCheminant, G.; Griffin, D.; Zeikus, R.; McIntosh, J.M.; Galyean, R.; Varga, J. Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using  $\omega$ -conotoxin from *Conus magus* venom. *Biochemistry* **1987**, *26*, 2086–2090. [CrossRef]
108. Brady, R.M.; Zhang, M.; Gable, R.; Norton, R.S.; Baell, J.B. De novo design and synthesis of a  $\mu$ -conotoxin KIIIA peptidomimetic. *Bioorganic Med. Chem. Lett.* **2013**, *23*, 4892–4895. Available online: <https://www.sciencedirect.com/science/article/pii/S0960894X13008196> (accessed on 26 December 2024). [CrossRef] [PubMed]
109. Duda, T.F.; Palumbi, S.R. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6820–6823. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC21999/> (accessed on 26 December 2024). [CrossRef] [PubMed]
110. Wang, H.; Li, X.; Zhangsun, D.; Yu, G.; Su, R.; Luo, S. The  $\alpha 9\alpha 10$  nicotinic acetylcholine receptor antagonist  $\alpha$ O-conotoxin GeXIVA [1,2] alleviates and reverses chemotherapy-induced neuropathic pain. *Mar. Drugs* **2019**, *17*, 265. Available online: <https://www.mdpi.com/1660-3397/17/5/265> (accessed on 26 December 2024). [CrossRef]
111. Romero, H.K.; Christensen, S.B.; Di Cesare Mannelli, L.; Gajewiak, J.; Ramachandra, R.; Elmslie, K.S.; Vetter, D.E.; Ghelardini, C.; Iadonato, S.P.; Mercado, J.L.; et al. Inhibition of  $\alpha 9\alpha 10$  nicotinic acetylcholine receptors prevents chemotherapy-induced neuropathic pain. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1825–E1832. Available online: <https://www.pnas.org/doi/10.1073/pnas.1621433114> (accessed on 26 December 2024). [CrossRef]
112. Zouridakis, M.; Papakyriakou, A.; Ivanov, I.A.; Kasheverov, I.E.; Tsetlin, V.; Tzartos, S.; Giastas, P. Crystal structure of the monomeric extracellular domain of  $\alpha 9$  nicotinic receptor subunit in complex with  $\alpha$ -conotoxin RgIA: Molecular dynamics insights into RgIA binding to  $\alpha 9\alpha 10$  nicotinic receptors. *Front. Pharmacol.* **2019**, *10*, 474. Available online: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2019.00474/full> (accessed on 26 December 2024). [CrossRef] [PubMed]
113. Smallwood, T.B.; Clark, R.J. Advances in venom peptide drug discovery: Where are we at and where are we heading? *Expert Opin. Drug Discov.* **2021**, *16*, 1163–1173. [CrossRef] [PubMed]
114. Safavi-Hemami, H.; Gajewiak, J.; Karanth, S.; Robinson, S.D.; Ueberheide, B.; Douglass, A.D.; Schlegel, A.; Imperial, J.S.; Watkins, M.; Bandyopadhyay, P.K.; et al. Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl. Acad. Sci.* **2015**, *112*, 1743–1748. Available online: <https://www.pnas.org/doi/full/10.1073/pnas.1423857112> (accessed on 26 December 2024). [CrossRef]
115. Robinson, S.D.; Safavi-Hemami, H. Insulin as a weapon. *Toxicon* **2016**, *123*, 56–61. Available online: <https://www.sciencedirect.com/science/article/pii/S0041010116303038> (accessed on 26 December 2024). [CrossRef] [PubMed]
116. Bayrhuber, M.; Vijayan, V.; Ferber, M.; Graf, R.; Korukottu, J.; Imperial, J.; Garrett, J.E.; Olivera, B.M.; Terlau, H.; Zweckstetter, M.; et al. Conkunitzin-S1 is the first member of a new kunitz-type neurotoxin family: Structural and Functional Characterization\*. *J. Biol. Chem.* **2005**, *280*, 23766–23770. Available online: <https://www.sciencedirect.com/science/article/pii/S0021925820674294> (accessed on 26 December 2024). [CrossRef] [PubMed]
117. Finol-Urdaneta, R.K.; Remedi, M.S.; Raasch, W.; Becker, S.; Clark, R.B.; Strüver, N.; Pavlov, E.; Nichols, C.G.; French, R.J.; Terlau, H. Block of Kv1.7 potassium currents increases glucose-stimulated insulin secretion. *EMBO Mol. Med.* **2012**, *4*, 424–434. Available online: <https://www.embopress.org/doi/full/10.1002/emmm.201200218> (accessed on 26 December 2024). [CrossRef] [PubMed]
118. Hemu, X.; Tam, J.P. Macrocyclic antimicrobial peptides engineered from  $\omega$ -conotoxin. *Curr. Pharm. Des.* **2017**, *23*, 2131–2138. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5470054/> (accessed on 26 December 2024). [CrossRef]
119. Wang, M.; Liao, Z.; Zhangsun, D.; Wu, Y.; Luo, S. Engineering enhanced antimicrobial properties in  $\alpha$ -conotoxin RgIA through D-type amino acid substitution and incorporation of lysine and leucine residues. *Molecules* **2024**, *29*, 1181. Available online: <https://www.mdpi.com/1420-3049/29/5/1181> (accessed on 26 December 2024). [CrossRef]
120. Mei, D.; Lin, Z.; Fu, J.; He, B.; Gao, W.; Ma, L.; Dai, W.; Zhang, H.; Wang, X.; Wang, J.; et al. The use of  $\alpha$ -conotoxin ImI to actualize the targeted delivery of paclitaxel micelles to  $\alpha 7$  nAChR-overexpressing breast cancer. *Biomaterials* **2015**, *42*, 52–65. Available online: <https://www.sciencedirect.com/science/article/pii/S0142961214012216> (accessed on 26 December 2024). [CrossRef]
121. Mei, D.; Zhao, L.; Chen, B.; Zhang, X.; Wang, X.; Yu, Z.; Ni, X.; Zhang, Q.  $\alpha$ -conotoxin ImI-modified polymeric micelles as potential nanocarriers for targeted docetaxel delivery to  $\alpha 7$ -nAChR overexpressed non-small cell lung cancer. *Drug Deliv.* **2018**, *25*, 493–503. [CrossRef]
122. Osipov, A.V.; Terpinskaya, T.I.; Yanchanka, T.; Balashevich, T.; Zhmak, M.N.; Tsetlin, V.I.; Utkin, Y.N.  $\alpha$ -conotoxins enhance both the in vivo suppression of ehrlich carcinoma growth and in vitro reduction in cell viability elicited by cyclooxygenase and lipoxigenase inhibitors. *Mar. Drugs* **2020**, *18*, 193. Available online: <https://www.mdpi.com/1660-3397/18/4/193> (accessed on 26 December 2024). [CrossRef]
123. Mondal, S.; Bhavna, R.; Mohan Babu, R.; Ramakumar, S. Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification. *J. Theor. Biol.* **2006**, *243*, 252–260. Available online: <https://www.sciencedirect.com/science/article/pii/S002251930600261X> (accessed on 26 December 2024). [CrossRef]



124. Zaki, N.; Sibai, F.; Campbell, P. Conotoxin protein classification using pairwise comparison and amino acid composition: Toxin-aam. In Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, Dublin, Ireland, 12–16 July 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 323–330, ISBN 978-1-4503-0557-0. Available online: <https://dl.acm.org/doi/10.1145/2001576.2001621> (accessed on 26 December 2024).
125. Yuan, L.-F.; Ding, C.; Guo, S.-H.; Ding, H.; Chen, W.; Lin, H. Prediction of the types of ion channel-targeted conotoxins based on radial basis function network. *Toxicol. Vitro*. **2013**, *27*, 852–856. Available online: <https://linkinghub.elsevier.com/retrieve/pii/S0887233312003621> (accessed on 26 December 2024). [CrossRef]
126. Ding, H.; Deng, E.-Z.; Yuan, L.-F.; Liu, L.; Lin, H.; Chen, W.; Chou, K.-C. iCTX-Type: A Sequence-Based Predictor for Identifying the Types of Conotoxins in Targeting Ion Channels. *BioMed Res. Int.* **2014**, *2014*, 286419. Available online: <http://www.hindawi.com/journals/bmri/2014/286419/> (accessed on 26 December 2024). [CrossRef]
127. Wu, Y.; Zheng, Y.; Tang, H. Identifying the Types of Ion Channel-Targeted Conotoxins by Incorporating New Properties of Residues into Pseudo Amino Acid Composition. *BioMed Res. Int.* **2016**, *2016*, 3981478. [CrossRef]
128. Zhang, L.; Zhang, C.; Gao, R.; Yang, R.; Song, Q. Using the SMOTE technique and hybrid features to predict the types of ion channel-targeted conotoxins. *J. Theor. Biol.* **2016**, *403*, 75–84. Available online: <https://linkinghub.elsevier.com/retrieve/pii/S0022519316300686> (accessed on 26 December 2024). [CrossRef]
129. Hofmann, K. Sensitive protein comparisons with profiles and hidden markov models. *Brief. Bioinform.* **2000**, *1*, 167–178. [CrossRef]
130. Nastou, K.C.; Tsousis, G.N.; Papandreou, N.C.; Hamodrakas, S.J. MBPpred: Proteome-wide detection of membrane lipid-binding proteins using profile hidden markov models. *Biochim. Et Biophys. Acta BBA Proteins Proteom.* **2016**, *1864*, 747–754. Available online: <https://www.sciencedirect.com/science/article/pii/S1570963916300577> (accessed on 26 December 2024). [CrossRef]
131. Huo, L.; Zhang, H.; Huo, X.; Yang, Y.; Li, X.; Yin, Y. pHMM-tree: Phylogeny of profile hidden markov models. *Bioinformatics* **2017**, *33*, 1093–1095. [CrossRef]
132. Feldbauer, R.; Gosch, L.; Lüftinger, L.; Hyden, P.; Flexer, A.; Rattei, T. DeepNOG: Fast and accurate protein orthologous group assignment. *Bioinformatics* **2021**, *36*, 5304–5312. [CrossRef]
133. Koua, D.; Brauer, A.; Laht, S.; Kaplinski, L.; Favreau, P.; Remm, M. ConoDicator: A tool for prediction of conopeptide superfamilies. *Nucleic Acids Res.* **2012**, *40*, W238–W241. [CrossRef] [PubMed]
134. Waris, M.; Ahmad, K.; Kabir, M.; Hayat, M. Identification of DNA binding proteins using evolutionary profiles position specific scoring matrix. *Neurocomputing* **2016**, *199*, 154–162. Available online: <https://www.sciencedirect.com/science/article/pii/S0925231216300455> (accessed on 26 December 2024). [CrossRef]
135. Liu, X.; Lu, Y.; Wang, L.; Geng, W.; Shi, X.; Zhang, X. RF-PSSM: A combination of rotation forest algorithm and position-specific scoring matrix for improved prediction of protein-protein interactions between hepatitis C virus and human. *Big Data Min. Anal.* **2023**, *6*, 21–31. Available online: <https://ieeexplore.ieee.org/document/9962955> (accessed on 26 December 2024). [CrossRef]
136. Ruan, X.; Xia, S.; Li, S.; Su, Z.; Yang, J. Hybrid framework for membrane protein type prediction based on the PSSM. *Sci. Rep.* **2024**, *14*, 17156. Available online: <https://www.nature.com/articles/s41598-024-68163-7> (accessed on 26 December 2024). [CrossRef]
137. Manavalan, B.; Lee, J. FRTpred: A novel approach for accurate prediction of protein folding rate and type. *Comput. Biol. Med.* **2022**, *149*, 105911. [CrossRef]
138. Nithyanandam, S.; Sangaraju, V.K.; Manavalan, B.; Lee, G. Computational prediction of protein folding rate using structural parameters and network centrality measures. *Comput. Biol. Med.* **2023**, *155*, 106436. [CrossRef] [PubMed]
139. Gowd, K.H.; Dewan, K.K.; Iengar, P.; Krishnan, K.S.; Balaram, P. Probing peptide libraries from *Conus achatinus* using mass spectrometry and cDNA sequencing: Identification of  $\delta$  and  $\omega$ -conotoxins. *J. Mass. Spectrom.* **2008**, *43*, 791–805. Available online: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/jms.1377> (accessed on 26 December 2024). [CrossRef] [PubMed]
140. Wang, X.F.; Wang, J.M.; Wang, X.L.; Zhang, Y. Predicting the Types of Ion Channel-Targeted Conotoxins Based on AVC-SVM Model. *BioMed Res. Int.* **2017**, *2017*, 2929807. Available online: <https://www.hindawi.com/journals/bmri/2017/2929807/> (accessed on 26 December 2024).
141. Zuo, Y.; Li, Y.; Chen, Y.; Li, G.; Yan, Z.; Yang, L. PseKRAAC: A flexible web server for generating pseudo K-tuple reduced amino acids composition. *Bioinformatics* **2017**, *33*, 122–124. [CrossRef] [PubMed]
142. Yang, S.; Liu, D.; Song, Y.; Liang, Y.; Yu, H.; Zuo, Y. Designing a structure-function alphabet of helix based on reduced amino acid clusters. *Arch. Biochem. Biophys.* **2024**, *754*, 109942. [CrossRef] [PubMed]
143. Zheng, L.; Liu, D.; Li, Y.A.; Yang, S.; Liang, Y.; Xing, Y.; Zuo, Y. RaacFold: A webserver for 3D visualization and analysis of protein structure by using reduced amino acid alphabets. *Nucleic Acids Res.* **2022**, *50*, W633–W638. [CrossRef]
144. Liang, Y.; Yang, S.; Zheng, L.; Wang, H.; Zhou, J.; Huang, S.; Yang, L.; Zuo, Y. Research progress of reduced amino acid alphabets in protein analysis and prediction. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 3503–3510. [CrossRef] [PubMed]
145. Zheng, L.; Liu, D.; Yang, W.; Yang, L.; Zuo, Y. RaacLogo: A new sequence logo generator by using reduced amino acid clusters. *Brief. Bioinform.* **2021**, *22*, bbaa096. [CrossRef] [PubMed]
146. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

147. Guo, Y.; Zhang, Z.; Tang, F. Feature selection with kernelized multi-class support vector machine. *Pattern Recognit.* **2021**, *117*, 107988. Available online: <https://www.sciencedirect.com/science/article/pii/S0031320321001758> (accessed on 26 December 2024). [CrossRef]
148. Liu, Q.; Chen, C.; Zhang, Y.; Hu, Z. Feature selection for support vector machines with RBF kernel. *Artif. Intell. Rev.* **2011**, *36*, 99–115. [CrossRef]
149. Madugula, S.S.; Pujar, P.; Nammi, B.; Wang, S.; Jayasinghe-Arachchige, V.M.; Pham, T.; Mashburn, D.; Artiles, M.; Liu, J. Identification of family-specific features in Cas9 and Cas12 proteins: A machine learning approach using complete protein feature spectrum. *J. Chem. Inf. Model.* **2024**, *64*, 4897–4911. [CrossRef] [PubMed]
150. Rahman, M.S.; Rahman, M.K.; Kaykobad, M.; Rahman, M.S. isGPT: An optimized model to identify sub-golgi protein types using SVM and random forest based feature selection. *Artif. Intell. Med.* **2018**, *84*, 90–100. Available online: <https://www.sciencedirect.com/science/article/pii/S0933365717305171> (accessed on 26 December 2024). [CrossRef] [PubMed]
151. Liu, S.; Shi, T.; Yu, J.; Li, R.; Lin, H.; Deng, K. Research on Bitter Peptides in the Field of Bioinformatics: A Comprehensive Review. *Int. J. Mol. Sci.* **2024**, *25*, 9844. Available online: <https://www.mdpi.com/1422-0067/25/18/9844> (accessed on 26 December 2024). [CrossRef]
152. Liu, B.; Gao, X.; Zhang, H. BioSeq-Analysis2.0: An updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. *Nucleic Acids Res.* **2019**, *47*, e127. [CrossRef]
153. Li, H.; Pang, Y.; Liu, B. BioSeq-BLM: A platform for analyzing DNA, RNA, and protein sequences based on biological language models. *Nucleic Acids Res.* **2021**, *49*, e129. [CrossRef] [PubMed]
154. Wei, Z.S.; Han, K.; Yang, J.Y.; Shen, H.B.; Yu, D.J. Protein–protein interaction sites prediction by ensembling SVM and sample-weighted random forests. *Neurocomputing* **2016**, *193*, 201–212. Available online: <https://www.sciencedirect.com/science/article/pii/S0925231216002022> (accessed on 26 December 2024). [CrossRef]
155. Jia, S.; Hu, X.; Sun, L. The comparison between random forest and support vector machine algorithm for predicting  $\beta$ -hairpin motifs in proteins. *Engineering* **2013**, *5*, 391. Available online: <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=40823&#abstract> (accessed on 26 December 2024). [CrossRef]
156. Chen, X.; Wang, C.C.; Yin, J.; You, Z.H. Novel human miRNA-disease association inference based on random forest. *Mol. Ther. Nucleic Acids* **2018**, *13*, 568–579. Available online: <https://www.sciencedirect.com/science/article/pii/S2162253118302762> (accessed on 26 December 2024). [CrossRef]
157. Wu, Q.W.; Xia, J.F.; Ni, J.C.; Zheng, C.H. GAERF: Predicting lncRNA-disease associations by graph auto-encoder and random forest. *Brief. Bioinform.* **2021**, *22*, bbaa391. [CrossRef] [PubMed]
158. Aybey, B.; Zhao, S.; Brors, B.; Staub, E. Immune cell type signature discovery and random forest classification for analysis of single cell gene expression datasets. *Front. Immunol.* **2023**, *14*, 1194745. Available online: <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2023.1194745/full> (accessed on 26 December 2024). [CrossRef] [PubMed]
159. Puillandre, N.; Duda, T.F.; Meyer, C.; Olivera, B.M.; Bouchet, P. One, four or 100 genera? A new classification of the cone snails. *J. Molluscan Stud.* **2014**, *81*, 1–23. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4541476/> (accessed on 26 December 2024). [CrossRef] [PubMed]
160. Naamati, G.; Askenazi, M.; Linial, M. ClanTox: A classifier of short animal toxins. *Nucleic Acids Res.* **2009**, *37*, W363–W368. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2703885/> (accessed on 26 December 2024). [CrossRef]
161. Gupta, S.; Kapoor, P.; Chaudhary, K.; Gautam, A.; Kumar, R.; Raghava, G.P.S. In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* **2013**, *8*, e73957. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3772798/> (accessed on 26 December 2024). [CrossRef]
162. Cole, T.J.; Brewer, M.S. TOXIFY: A deep learning approach to classify animal venom proteins. *PeerJ* **2019**, *7*, e7200. Available online: <https://peerj.com/articles/7200> (accessed on 26 December 2024). [CrossRef]
163. Wheeler, T.J.; Eddy, S.R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **2013**, *29*, 2487–2489. [CrossRef] [PubMed]
164. Conticello, S.G.; Kowalsman, N.D.; Jacobsen, C.; Yudkovsky, G.; Sato, K.; Elazar, Z.; Petersen, C.M.; Aronheim, A.; Fainzilber, M. The prodomain of a secreted hydrophobic mini-protein facilitates its export from the endoplasmic reticulum by hitchhiking on sorting receptors. *J. Biol. Chem.* **2003**, *278*, 26311–26314. [CrossRef]
165. Buczek, O.; Olivera, B.M.; Bulaj, G. Propeptide does not act as an intramolecular chaperone but facilitates protein disulfide isomerase-assisted folding of a conotoxin precursor. *Biochemistry* **2004**, *43*, 1093–1101. [CrossRef]
166. Yu, H.F.; Huang, F.L.; Lin, C.J. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach. Learn.* **2011**, *85*, 41–75. [CrossRef]
167. Heerdt, G.; Zanutto, L.; Souza, P.C.T.; Araujo, G.; Skaf, M.S. Collision cross section calculations using HPCCS. *Methods Mol. Biol.* **2020**, *2084*, 297–310. [PubMed]

168. Chi, Q.N.; Jia, S.X.; Yin, H.; Wang, L.E.; Fu, X.Y.; Ma, Y.N.; Sun, M.P.; Qi, Y.K.; Li, Z.B.; Du, S.S. Efficient synthesis and anticancer evaluation of spider toxin peptide LVTX-8-based analogues with enhanced stability. *Bioorganic Chem.* **2023**, *134*, 106451. Available online: <https://www.sciencedirect.com/science/article/pii/S0045206823001116> (accessed on 26 December 2024). [CrossRef] [PubMed]
169. Greco, I.; Molchanova, N.; Holmedal, E.; Jenssen, H.; Hummel, B.D.; Watts, J.L.; Håkansson, J.; Hansen, P.R.; Svenson, J. Correlation between hemolytic activity, cytotoxicity and systemic in vivo toxicity of synthetic antimicrobial peptides. *Sci. Rep.* **2020**, *10*, 13206. Available online: <https://www.nature.com/articles/s41598-020-69995-9> (accessed on 26 December 2024). [CrossRef]
170. Robinson, S.D.; Undheim, E.A.B.; Ueberheide, B.; King, G.F. Venom peptides as therapeutics: Advances, challenges and the future of venom-peptide discovery. *Expert Rev. Proteom.* **2017**, *14*, 931–939. Available online: <https://www.tandfonline.com/doi/abs/10.1080/14789450.2017.1377613> (accessed on 26 December 2024). [CrossRef] [PubMed]
171. Spotlight on researchers during a wet-lab to dry-lab transition period: An interview with guifen liu and qi wang. *Commun. Biol.* **2024**, *7*, 1–3. Available online: <https://www.nature.com/articles/s42003-024-06923-x> (accessed on 26 December 2024).
172. Asamitsu, S.; Yabuki, Y.; Matsuo, K.; Kawasaki, M.; Hirose, Y.; Kashiwazaki, G.; Chandran, A.; Bando, T.; Wang, D.O.; Sugiyama, H.; et al. RNA G-quadruplex organizes stress granule assembly through DNAPTP6 in neurons. *Sci. Adv.* **2023**, *9*, eade2035. Available online: <https://www.science.org/doi/10.1126/sciadv.ade2035> (accessed on 26 December 2024). [CrossRef]
173. Madani, A.; Krause, B.; Greene, E.R.; Subramanian, S.; Mohr, B.P.; Holton, J.M.; Olmos, J.L.; Xiong, C.; Sun, Z.Z.; Socher, R.; et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, *41*, 1099–1106. Available online: <https://www.nature.com/articles/s41587-022-01618-2> (accessed on 26 December 2024). [CrossRef] [PubMed]
174. Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **2022**, *13*, 4348. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9329459/> (accessed on 26 December 2024). [CrossRef] [PubMed]
175. Sumida, K.H.; Núñez-Franco, R.; Kalvet, I.; Pellock, S.J.; Wicky, B.I.M.; Milles, L.F.; Dauparas, J.; Wang, J.; Kipnis, Y.; Jameson, N.; et al. Improving protein expression, stability, and function with ProteinMPNN. *J. Am. Chem. Soc.* **2024**, *146*, 2054–2061. [CrossRef]
176. Guo, M.; Li, Z.; Deng, X.; Luo, D.; Yang, J.; Xue, W. ConoDL: A Deep Learning Framework for Rapid Generation and Prediction of Conotoxins. *J. Comput. Mol. Des.* **2024**, *39*, 1–15. [CrossRef] [PubMed]
177. Tay, Y.; Dehghani, M.; Bahri, D.; Metzler, D. Efficient transformers: A survey. *ACM Comput. Surv.* **2022**, *55*, 109:1–109:28. Available online: <https://dl.acm.org/doi/10.1145/3530811> (accessed on 26 December 2024). [CrossRef]
178. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.U.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30. Available online: [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html) (accessed on 26 December 2024).
179. Yang, L.; Yang, G.; Bing, Z.; Tian, Y.; Huang, L.; Niu, Y.; Yang, L. Accelerating the discovery of anticancer peptides targeting lung and breast cancers with the wasserstein autoencoder model and PSO algorithm. *Brief. Bioinform.* **2022**, *23*, bbac320. [CrossRef]
180. Cao, Q.; Ge, C.; Wang, X.; Harvey, P.J.; Zhang, Z.; Ma, Y.; Wang, X.; Jia, X.; Mobli, M.; Craik, D.J.; et al. Designing antimicrobial peptides using deep learning and molecular dynamic simulations. *Brief. Bioinform.* **2023**, *24*, bbad058. Available online: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbad058/7066348> (accessed on 26 December 2024). [CrossRef]
181. Das, P.; Sercu, T.; Wadhawan, K.; Padhi, I.; Gehrmann, S.; Cipcigan, F.; Chenthamarakshan, V.; Strobel, H.; dos Santos, C.; Chen, P.Y.; et al. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nat. Biomed. Eng.* **2021**, *5*, 613–623. Available online: <https://www.nature.com/articles/s41551-021-00689-x> (accessed on 26 December 2024). [CrossRef] [PubMed]
182. Ali, F.; Khalid, M.; Almuhaimeed, A.; Masmoudi, A.; Alghamdi, W.; Yafoz, A. IP-GCN: A deep learning model for prediction of insulin using graph convolutional network for diabetes drug design. *J. Comput. Sci.* **2024**, *81*, 102388. [CrossRef]
183. Ali, F.; Khalid, M.; Masmoudi, A.; Alghamdi, W.; Yafoz, A.; Alsini, R. VEGF-ERCNN: A Deep Learning-based Model for Prediction of Vascular Endothelial Growth Factor using Ensemble Residual CNN. *J. Comput. Sci.* **2024**, *83*, 102448. [CrossRef]
184. Ali, F.; Almuhaimeed, A.; Khalid, M.; Alshanbari, H.; Masmoudi, A.; Alsini, R.J.M. DEEP-EP: Identification of epigenetic protein by ensemble residual convolutional neural network for drug discovery. *Methods* **2024**, *226*, 49–53. [CrossRef] [PubMed]
185. Alsini, R.; Almuhaimeed, A.; Ali, F.; Khalid, M.; Farrash, M.; Masmoudi, A. Deep-VEGF: Deep stacked ensemble model for prediction of vascular endothelial growth factor by concatenating gated recurrent unit with two-dimensional convolutional neural network. *J. Biomol. Struct. Dyn.* **2024**, 1–11. [CrossRef]
186. Sohl-Dickstein, J.; Weiss, E.A.; Maheswaranathan, N.; Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. *arXiv* **2015**, arXiv:1503.03585v8. Available online: <https://arxiv.org/abs/1503.03585v8> (accessed on 26 December 2024).
187. Cai, Z.; Xiong, Z.; Xu, H.; Wang, P.; Li, W.; Pan, Y. Generative adversarial networks: A survey toward private and secure applications. *ACM Comput. Surv.* **2021**, *54*, 132:1–132:38. Available online: <https://dl.acm.org/doi/10.1145/3459992> (accessed on 26 December 2024). [CrossRef]



188. Sohn, K.; Yan, X.; Lee, H. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems 28 (Nips 2015)*; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; Neural Information Processing Systems (nips): La Jolla, CA, USA, 2015; Volume 28. Available online: <https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000450913101035> (accessed on 26 December 2024).
189. Lu, J.; Batra, D.; Parikh, D.; Lee, S. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In *Advances in Neural Information Processing Systems 32 (Nips 2019)*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alche-Buc, F., Fox, E., Garnett, R., Eds.; Neural Information Processing Systems (nips): La Jolla, CA, USA, 2019; Volume 32. Available online: <https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000534424300002> (accessed on 26 December 2024).
190. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. Available online: <https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000262429700005> (accessed on 26 December 2024). [CrossRef]
191. Gao, Y.; Chuai, G.; Yu, W.; Qu, S.; Liu, Q. Data imbalance in CRISPR off-target prediction. *Brief. Bioinform.* **2020**, *21*, 1448–1454. [CrossRef]
192. Tang, X.; Cai, L.; Meng, Y.; Gu, C.; Yang, J.; Yang, J. A novel hybrid feature selection and ensemble learning framework for unbalanced cancer data diagnosis with transcriptome and functional proteomic. *IEEE Access* **2021**, *9*, 51659–51668. Available online: <https://ieeexplore.ieee.org/document/9393342> (accessed on 26 December 2024). [CrossRef]
193. Pham, N.T.; Phan, L.T.; Seo, J.; Kim, Y.; Song, M.; Lee, S.; Jeon, Y.J.; Manavalan, B. Advancing the accuracy of SARS-CoV-2 phosphorylation site detection via meta-learning approach. *Brief. Bioinform.* **2023**, *25*, bbad433. [CrossRef]
194. Pham, N.T.; Zhang, Y.; Rakkiyappan, R.; Manavalan, B. HOTGpred: Enhancing human O-linked threonine glycosylation prediction using integrated pretrained protein language model-based features and multi-stage feature selection approach. *Comput. Biol. Med.* **2024**, *179*, 108859. [CrossRef] [PubMed]
195. Joshi, M.; Singh, B.K. Deep Learning Techniques for Brain Lesion Classification Using Various MRI (from 2010 to 2022): Review and Challenges. *Medinformatics* **2024**, 1–21. [CrossRef]
196. Zulfikar, H.; Guo, Z.; Ahmad, R.M.; Ahmed, Z.; Cai, P.; Chen, X.; Zhang, Y.; Lin, H.; Shi, Z. Deep-STP: A deep learning-based approach to predict snake toxin proteins by using word embeddings. *Front. Med.* **2024**, *10*, 1291352. [CrossRef] [PubMed]
197. Wang, H.; Li, J.; Wu, H.; Hovy, E.; Sun, Y. Pre-trained language models and their applications. *Engineering* **2023**, *25*, 51–65. Available online: <https://www.sciencedirect.com/science/article/pii/S2095809922006324> (accessed on 26 December 2024). [CrossRef]
198. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2021**, *109*, 43–76. Available online: <https://ieeexplore.ieee.org/document/9134370> (accessed on 26 December 2024). [CrossRef]
199. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
200. Charuvaka, A.; Rangwala, H. Classifying protein sequences using regularized multi-task learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2014**, *11*, 1087–1098. Available online: <https://ieeexplore.ieee.org/document/6855357> (accessed on 26 December 2024). [CrossRef] [PubMed]
201. Lin, S.; Shi, C.; Chen, J. GeneralizedDTA: Combining pre-training and multi-task learning to predict drug-target binding affinity for unknown drug discovery. *BMC Bioinform.* **2022**, *23*, 367. [CrossRef] [PubMed]
202. Yuan, Q.; Chen, S.; Wang, Y.; Zhao, H.; Yang, Y. Alignment-free metal ion-binding site prediction from protein sequence through pretrained language model and multi-task learning. *Brief. Bioinform.* **2022**, *23*, bbac444. [CrossRef]
203. Mozhaeva, V.; Kudryavtsev, D.; Prokhorov, K.; Utkin, Y.; Gudkov, S.; Garnov, S.; Kasheverov, I.; Tsetlin, V. Toxins' classification through raman spectroscopy with principal component analysis. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *278*, 121276. Available online: <https://www.sciencedirect.com/science/article/pii/S1386142522004255> (accessed on 26 December 2024). [CrossRef]
204. Zou, X.; Ren, L.; Cai, P.; Zhang, Y.; Ding, H.; Deng, K.; Yu, X.; Lin, H.; Huang, C. Accurately identifying hemagglutinin using sequence information and machine learning methods. *Front. Med.* **2023**, *10*, 1281880. [CrossRef]
205. Fujita, T. In memoriam professor corwin hansch: Birth pangs of QSAR before 1961. *J. Comput. Aided Mol. Des.* **2011**, *25*, 509–517. [CrossRef]
206. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.; Agrawal, R. Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov.* **2011**, *3*, 511–519. Available online: <https://www.semanticscholar.org/paper/Validation-of-QSAR-Models-Strategies-and-Importance-Veerasamy-Rajak/4eb25ff5a87f2fd6789c5b9954eddddf1c59dab> (accessed on 26 December 2024). [CrossRef]
207. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476–488. Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1002/minf.201000061> (accessed on 26 December 2024). [CrossRef]
208. He, S.; Ye, X.; Sakurai, T. CLOP-hERG: The Contrastive Learning Optimized Pre-Trained Model for Representation Learning in Predicting Drug-Induced hERG Channel Blockers. *Medinformatics* **2024**, *1*, 103–111. [CrossRef]

- 
209. Xiang, H.; Zeng, L.; Hou, L.; Li, K.; Fu, Z.; Qiu, Y.; Nussinov, R.; Hu, J.; Rosen-Zvi, M.; Zeng, X.J.N.C. A molecular video-derived foundation model for scientific drug discovery. *Nat. Commun.* **2024**, *15*, 9696. [[CrossRef](#)]
210. Li, T.; Ren, X.; Luo, X.; Wang, Z.; Li, Z.; Luo, X.; Shen, J.; Li, Y.; Yuan, D.; Nussinov, R.J.N.C. A foundation model identifies broad-spectrum antimicrobial peptides against drug-resistant bacterial infection. *Nat. Commun.* **2024**, *15*, 7538. [[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.