



Review

# Venomics-Accelerated Cone Snail Venom Peptide Discovery

S. W. A. Himaya and Richard J. Lewis \*

IMB Centre for Pain Research, Institute for Molecular Bioscience, The University of Queensland, Queensland 4072, Australia; h.siddhialu@imb.uq.edu.au

\* Correspondence: r.lewis@imb.uq.edu.au; Tel.: +61-7-3346-2984; Fax: +61-7-3346-2101

Received: 18 January 2018; Accepted: 21 February 2018; Published: 9 March 2018

**Abstract:** Cone snail venoms are considered a treasure trove of bioactive peptides. Despite over 800 species of cone snails being known, each producing over 1000 venom peptides, only about 150 unique venom peptides are structurally and functionally characterized. To overcome the limitations of the traditional low-throughput bio-discovery approaches, multi-omics systems approaches have been introduced to accelerate venom peptide discovery and characterisation. This “venomic” approach is starting to unravel the full complexity of cone snail venoms and to provide new insights into their biology and evolution. The main challenge for venomics is the effective integration of transcriptomics, proteomics, and pharmacological data and the efficient analysis of big datasets. Novel database search tools and visualisation techniques are now being introduced that facilitate data exploration, with ongoing advances in related omics fields being expected to further enhance venomics studies. Despite these challenges and future opportunities, cone snail venomics has already exponentially expanded the number of novel venom peptide sequences identified from the species investigated, although most novel conotoxins remain to be pharmacologically characterised. Therefore, efficient high-throughput peptide production systems and/or banks of miniaturized discovery assays are required to overcome this bottleneck and thus enhance cone snail venom bioprospecting and accelerate the identification of novel drug leads.

**Keywords:** cone snails; venom; venomics; proteomics; transcriptomics; visualisation

## 1. Introduction

Millions of years of evolution have generated natural products that can be successfully used as drugs and drug candidates, with >50% of all approved drugs arising from natural products or their derivatives [1]. Although plant-based natural products have been more popular as leads, animal venoms are also regarded as a treasure trove of potential therapeutics [2,3]. Since the success of Captopril, an inhibitor of angiotensin-converting enzyme (ACE) identified from *Bothrops jararaca* snake venom in 1981, drug discovery research using animal venoms has gained momentum [3,4]. Marine cone snails have emerged as one of the most promising sources of new drug leads. This was first shown in 2004 with the FDA approval of Zinconotide, a synthetic form of the  $\omega$ -conotoxin MVIIA isolated from *Conus magus* used to treat intractable pain [5]. More recently, other conotoxins, including Xen2174 (Mr1A), a norepinephrine transporter antagonist isolated from the *Conus marmoreus* [6], CGX-1007 (Conantokin G), a *N*-methyl-D-aspartate (NMDA) receptor antagonist from *Conus geographus* [7], CGX-1051, a potassium channel blocker from *Conus purpurascens* [8], and ACV1 (Vc1.1) an AChR antagonist identified from *Conus victoriae* [9], have also progressed into the clinic, although only Xen2174 proved efficacious in patients. To enhance the search for successful drug candidates, high-throughput discovery assays are being used in combination with proteomic and transcriptomic approaches and sophisticated structure–activity analyses to accelerate the hit-to-lead process.

Cone snails are marine predators that hunt fish, mollusks, and worms depending on their prey preference, using sophisticated envenomation strategies for prey capture and defense. The ~800 species of cone snails contain a rich variety of small cysteine-rich peptides called conotoxins, together with a lesser number of cysteine-poor conopeptides with diverse pharmacology, that target prey, predator, and/or mammalian receptors with high affinity and specificity [2,10,11]. Cone snail venoms are highly complex with layers of diversity triggered by dietary specialization [12,13] and predatory or defensive stimuli. It has been found that each cone snail species produces a unique venom comprising >1000 venom components [14,15]. Adding to this complexity, variation across individuals of the same species further expands the pool of potentially bioactive peptides available [16–22].

As drug candidates, peptides possess inherent advantages of higher selectivity and affinity with fewer off-target interactions compared to small molecules, with the promise of better efficacy with potentially lower side effects [23]. Moreover, advanced bio-engineering approaches are now being used to improve the stability and oral bioavailability of peptides [24,25]. In this context, cone snail venom peptides present an ideal source for the mining of potential drug candidates. However, the expected growth in discovery has often been limited by low-throughput, scarcely sensitive discovery methods and material limitations. Typically, abundant peptides from common species such as *C. geographus*, *Conus striatus*, *C. magus*, *Conus catus*, *C. purpurascens*, and *Conus textiles* are most comprehensively studied, while the large pool of minor venom components remain mostly uncharacterized. Tapping into this peptide reserve requires a broader implementation of miniaturised high-throughput multidimensional strategies. This review provides an overview of the impact of these technological advancements on cone snail venom research over the past decade and discusses the challenges and opportunities that remain.

## 2. Venomics: Heralding a New Era of Venom Bioprospecting

Despite cone snail venom peptides being highly regarded as potential drug leads and the diversity and large size of the peptide pool, less than 150 unique peptides have been structurally and pharmacologically characterized over the past 30 years, representing significantly less than 1% of the total conopeptide diversity. To unravel the complex pharmacology of conotoxins and to isolate interesting minor active components using a single venom is challenging as it requires significant quantities of crude venom. Sometimes, tens to hundreds of cone snails are necessary to obtain the venom required to characterize a single molecule of interest [26]. This practice can be a waste of a precious resource and warrants the implementation of miniaturized and accelerated approaches that can better support sustainable bioprospecting.

In recent years, parallel developments in mass spectrometry [27–29], next-generation sequencing, [30–32] and high-throughput miniaturized target screens [33] have largely benefited venom researchers. For example, adaptation of multi-omics systems approaches to cone snail venom discovery pipelines have enhanced the bio-discovery process and improved our understanding of venom peptide evolution and origin using a small number of animals [15]. By their nature, non-targeted and non-biased transcriptomics and proteomics approaches provide a holistic view of the components that make up cone snail venoms. The integration of these omics approaches to study venoms was termed “venomics” [34]. Since its implementation, cone snail venom research has advanced exponentially with the support of specialized databases (conoserver) [10,11] and automated (conosorter, conodictor) [35,36] and semi-automated analysis pipelines [37] and knowledge bases. Using this integrative technology, full repertoires of venoms were revealed from even rarely studied species, and novel peptide classes and frameworks have been discovered that provide new opportunities for bioprospecting. Perhaps even more interesting, this multi-dimensional approach has helped generate novel hypotheses on venom evolution [14,15,38,39] and cone snail behavior and prey capture strategies [14,20,40,41].

### 2.1. Cone Snail Venom Duct Transcriptomics

In cone snails, the apical secretory cells lining the long, convoluted venom duct [42,43] translates mature mRNA to precursor conopeptides that typically comprise three distinct regions: an N-terminal

endoplasmic reticulum (ER) signal sequence, a central pro-peptide region, and the C-terminal mature toxin. Cone snail venom duct transcriptomes reveal the transcripts of all expressed proteins and peptides, including minor peptides and rare variants, providing the template for active venom peptide synthesis during translation [44]. In 2007, the CONCO cone snail venomomics project was launched with twenty partnering companies and universities, with the aim of applying this novel approach for the accelerated, cheaper, safer, and more ethical production of innovative biomedical drugs from *Conus consors* and related species [36]. Since then, complete venom transcript repertoires of over a dozen cone snail species have been published using next-generation sequencing platforms such as 454 (Roche, Branford, CT, USA), Illumina (Illumina, San Diego, CA, USA), Ion Torrent™ Personal Genome Machine™ (PGM) (Thermo Fisher, Waltham, MA, USA), Nanopore (Oxford, UK), ABI 3730 Series (Applied Biosystems, Foster City, CA, USA) and PacBio (Pacific Biosciences, Menlo Park, CA, USA) (Table 1).

**Table 1.** Performance of next-generation sequencing platforms.

Developer	System	Read Length (bp)	Error Rate (%)
Illumina	Illumina Hi-Seq2000	2 × 150 bp	0.1
	Illumina Hi-Seq3000/4000	2 × 150 bp	0.1
	Illumina Mi-Seq (Version 3)	2 × 300 bp	0.1
Thermo Fisher Scientific	Ion Torrent PGM 318	400 bp	1
Applied Biosystems	ABI 3730/ABI 3730xl	500 bp	0.4
Pacific Biosciences	PacBio RS II	10–20 kb	10–15
	PacBio Sequel	>20 kb	10–15
Oxford	Nanopore PromethION	900 kb	4
Roche	454 (discontinued in 2013)	700 bp	1

Initially, the 454 (Roche) pyrosequencing platform was used to obtain the cone snail venom duct transcriptomes [14,15,20,38,40]. However, with the discontinuation of 454 technology in 2013, Illumina is now most widely used to generate venom duct transcriptomes. The transcriptomic data quality is determined by the RNA sequence data read length and the sequencing error rate [45]. Transcriptomes generated through the 454 platform contained longer reads (700 bp single-end) than the Illumina systems (150–300 bp paired-end), though the error rate of Illumina is significantly lower than that of the 454 platform (Table 1). The shorter reads generated in Illumina platforms require data assembly [46] using carefully optimized parameters and validation to remove artifacts [44]. This is especially true for cone snail venom duct transcriptome assembly where no reference genome is available as a guide. In contrast, the PacBio system generates longer reads but with a higher error rate that makes this system a poor fit for cone snail transcriptome studies, especially when transcriptomic messiness is being investigated [38]. Recently, the complete venom duct transcriptome of *Conus betulinus* [47] ABI 3730 was generated from intermediate length reads (500 bp) with low error rates. As the average length of a conotoxin precursor is ~70 amino acids (~210 nucleotides), platforms offering above 2 × 300 bp nucleotide read coverage with low error rates allow transcriptomes to be faithfully generated.

Irrespective of the platform used, next-generation RNA sequencing generates large data sets that require purpose-built, dedicated bioinformatics tools for efficient data mining. ConoSorter (version 1.1, The University of Queensland, Australia) is one of the most successful standalone tool for rapid and accurate interrogation of thousands of sequences produced by high-throughput sequencing methods. ConoSorter employs Hidden Markov Models for putative toxin identification and annotation. The tool categorizes cDNA or protein sequences into conopeptide superfamilies and classes on the basis of their signal, propeptide, and mature regions. ConoSorter also facilitates the cataloging of the main sequence characteristics (relative sequence frequency, length, number of cysteines, N-terminal hydrophobicity, sequence similarity score) and automatically searches the ConoServer (The University of Queensland, Australia) database of all described conopeptides for known precursor sequences, enabling the rapid identification of known and novel conopeptides. It was reported that, when applied to ConoServer

and UniProtKB/Swiss-Prot databases, ConoSorter recognized 100% of known conotoxin superfamilies and classes with a minimum species specificity of 99%, and thus it remains the best publicly available tool for cone snail venom duct transcriptome mining [35].

The growing number of transcriptomic projects have already revealed a wealth of new knowledge about the conopeptides, including the discovery of novel superfamilies and frameworks (Table 2), and have helped uncover novel peptide pharmacology [41,48–51]. Using this approach, the messy processing at the transcriptomic level was identified as a significant generator of venom peptide diversity [38]. Importantly, messiness is considered a novel mechanism of adaptation in cone snails that contributes to the rapid evolution of venom peptides with new functions.

**Table 2.** New superfamilies discovered by transcriptomics.

Species	New Superfamily	Cysteine Framework	Sequencing	Reference
<i>Conus arenatus</i> <i>Conus coronatus</i> <i>Conus rattus</i>	MRFYM-	VI/VII	Illumina HiSeq 2000	[13]
<i>Conus arenatus</i> <i>Conus imperialis</i> <i>Conus lividus</i> <i>Conus quercinus</i> <i>Conus sponsalis</i> <i>Conus virgo</i>	MKISL-	VI/VII	Illumina HiSeq 2000	[13]
<i>Conus marmoreus</i>	N, B, H, E, F, H2, I4, M2, N2, O4, Q, R, U, W, X, Y2, Y3, Z	XV, VIII, VI/VII, N, N, III, C1, C7*, N, VI/VII, N, N, VI/VII, N, C2, C4*, C2, C2	Roche 454	[15]
<i>Conus catus</i>	Cat-NSF1	VI/VII	Roche 454	[20]
<i>Conus miles</i>	SF-mi 1–8	XIII, C8-novel, VI/VII, N, N, XV, NA, XIII	Roche 454	[38]
<i>Conus vexillum</i>	NSVx1-4	XIV, XXIV, VI/VII, XXVIII (novel)	Roche 454	[40]
<i>Conus betulinus</i>	NSF-bt01–09	IX, XV, VI/VII, XIV, VI/VII, IX, VIII, VI/VII, IX	ABI 3730	[47]
<i>Conus episcopatus</i>	SF-Epi 1–16	variable, III <sup>a</sup> , V, C7*, V, V, V, XI, C7*, N, C5*, NA, XVIII, VI/VII, IV, C5*	Illumina MiSeq	[48]
<i>Conus tribblei</i> and <i>Conus lenavati</i>	SF-01-04	C12*, C12*, IX, XIII,	Illumina HiSeq 2000	[49]
<i>Conus pullicarius</i>	J2	XIV	Roche GS-FLX	[52]
<i>Conus gloriamaris</i>	MKAVA-, MSRLF-, MMLFM-, MLSML-	XXII, N, VIII, C12*	Illumina HiSeq 2000	[53]
<i>Conus lenavati</i>	SF-05-06	XIV, C12*	Illumina HiSeq 2000	[54]
<i>Conus andremenezi</i>	Put.MGGRE, Put.MKAVA	N, C8*	Illumina HiSeq 2000	[55]
<i>Conus andremenezi</i> <i>Conus praececellens</i>	Put.MSGLR, Put.MUSGK	VI/VII, C6*	Illumina HiSeq 2000	[55]

N, no cysteines were detected in the mature sequence; \* no framework name has been defined, possibly a novel cysteine framework; <sup>a</sup> probable predicted framework.

## 2.2. Proteomics—Exploring the Complexity of Expressed Venom Peptides

The term “proteomics” was first coined in 1995 to define the large-scale characterization of the entire protein profile of a cell line, tissue, or organism [28,56] that underpins the search for a global and integrated view of biology by understanding all proteins and their interactions [57]. Advances in the field of proteomics and the ongoing developments in state-of-the-art instrumentation has greatly benefitted cone snail venom research. Traditional venom peptide identification requires individual peptide isolation by HPLC followed by Edman degradation to identify sequences of individual peptides [15]. Unfortunately, these techniques consume considerable amount of material and are time-consuming. Proteomics provides a versatile approach that supports a comprehensive,

multi-pronged analysis of venom peptides, including sequence, quantity, post translational modifications, regionalisation, and stimulus-dependence of venom peptide mobilisation. In cone snail research, proteomics accommodates two main research approaches, i.e., the rapid identification of specific classes of peptides of interest and the systematic big data analysis for comprehensive venom studies. Large-scale proteomic approaches integrated with matching transcriptomes can help generate novel biological insights by providing a global view of venom complexity beyond the capabilities of classical venom peptide characterization approaches. Understandably, these mass spectrometric methods generate a wealth of data that require comprehensive reference databases and dedicated computational tools to fully exploit their potential.

### 2.3. Mass Spectrometry in Venom Characterisation

Mass spectrometric methods have enabled the characterisation of thousands of different peptides at widely varying concentrations [58] in complex mixtures such as venoms. Liquid chromatography (LC) coupled to mass spectrometry (MS) or tandem mass spectrometry (MS/MS) underpin such proteomic studies [59]. Because of the complex nature of cone snail venoms, it is critical to obtain high-resolution MS data under optimized HPLC conditions to separate and decipher these complex mixtures in a single run. Ultra-high-performance liquid chromatography (uHPLC) and, more recently, microfluidic techniques such as nano-liquid chromatography (Nano-LC) and capillary electro-chromatography (CEC) can deliver remarkable separations using low sample amounts (>2 µg) and flow rates (<1 µL/min) [60]. These techniques provide high selectivity, mass detection sensitivity, separation efficiency, resolution, and rapid analysis and are ideal for the comprehensive proteomics analysis of venoms.

For cone snail proteomic studies, Data-Dependent Acquisition (DDA)-based or Information-Dependent Acquisition (IDA)-based mass spectrometric detection methods are commonly used [15,20,38,48]. The DDA mode selects predetermined  $m/z$  values of interest for secondary MS/MS analysis, providing better ionization of the targeted peptides of interest. IDA serves as an artificial intelligence-based scan mode which provides “on-the-fly” acquisition of MS/MS spectra during a HPLC run to obtain specific, selective, and information rich MS/MS from precursor ions [55]. Not surprisingly, these methods have greatly facilitated the generation of MS/MS lists of peptides from crude cone snail venoms for database searches and sequence validation. Time-of-flight (TOF) mass spectrometry, especially quadrupole–time-of-flight (Q-TOF) MS, provides fast acquisition speeds, superior sensitivity, high resolution, and excellent mass accuracy [43], characteristics that are all required for venom studies. For example, by using these state-of-the-art techniques it was found that each crude venom comprises >1000 peptides, while traditional peptide detection methods typically found only ~100 peptides [15,20,38,48].

Comprehensive MS/MS libraries of venom peptide fragments facilitate the integration with venom duct transcriptomic data, allowing the relatively straightforward assignment of full sequences for many of the peptides present in the proteome, in a bottom-up proteomics approach [58]. Typically, cone snail venoms are digested into peptide fragments by a sequence-specific enzyme such as trypsin or Glu C, and MS/MS spectra (IDA- or DDA-assisted) are generated that contain sufficient sequence information to identify and quantify the parent peptides [15,20,38]. Collision-induced dissociation or higher-energy collisional dissociation [61] methods are often used for ion fragmentation. However, alternative methods such as electron transfer dissociation [62] are becoming more widely available and promise improved fragmentation of larger and modified peptides. These novel methods can be readily incorporated into the cone snail venom peptide identification pipeline for high-throughput peptide identification and validation.

### 2.4. Omics Data Integration

The use of omics data to interpret complex proteinaceous compositions became possible with the introduction of the comprehensive databases and efficient database searching algorithms [57].



These databases and search tools enable the integration of transcriptomic data into the proteomic data and into a powerful pipeline for rapid peptide identification and quantification. However, there are few studies that reveal close to the full set of toxins produced in the venom gland of even a single species. To date, the deep-venomics study of *C. marmoreus* still remains the most comprehensive, with 60% of the transcriptome (105 precursors) explained by 1385 peptide fragments sequenced by MS/MS [15,35]. Despite the challenges in finding all major transcriptomic sequences in the proteome, venomics remains the fastest and most efficient approach to obtain broad-scale sequence data, even for highly divergent peptides of minor gene superfamilies [15,20,35,38,40,49]. The major bottleneck in venom data analysis and interpretation is the lack of dedicated reference databases and robust data analysis tools that can integrate all aspects of conotoxin diversity and complexity. Currently, UniProt, Conoserver, and venom transcriptome sequence libraries serve as reference databases for proteomic search, with software commonly used to validate transcriptomic sequences against tandem mass spectrometric data listed in Table 3. As the field of proteomics continues to advance, novel high-resolution mass spectrometry in combination with new dedicated venom databases and optimized search engines are expected to improve the sequence recovery rates. User-friendliness, reliable post-translational modification (PTM) prediction and scoring, and improved data visualisation tools are expected to facilitate venom peptide proteome and transcriptome integration and help unravel the underlying biology driving their remarkable diversity.

**Table 3.** Commonly used search algorithms and proteomic analysis tools. PTM: post-translational modification.

Tool	Function
<b>Database search algorithms</b>	
PEAKS DB	Database search engine, run in parallel with de novo sequencing, to automatically validate the search results
ProteinPilot Software	Enables peptide identification while considering PTMs, non-tryptic cleavages, and amino acid substitutions. Supports quantification using iTRAQ, mTRAQ, and SILAC
Mascot	Peptide mass fingerprinting and MS/MS database searching
Protein Prospector	Proteomic analysis tools including “Batch-Tag” for instrument- and fragmentation mode-optimised analysis
MassMatrix	Search algorithm for tandem MS data that ranks peptide and protein matches
Byonic	MS/MS data search for fragment identifications to produce protein scores and identification probabilities
<b>MS/MS peptide quantification and visualisation software</b>	
MarkerView Software	Statistical analysis and visualisation of quantitative mass spec data sets from proteomic profiling applications
MultiQuant Software	Quantitation and targeted visualisation of TripleTOF or QTRAP data, including MRM and SWATH acquisition
MaxQuant	Quantitative proteomics tool for analysis of label-free and SILAC-based proteomics data

### 3. Complex Venom Processing Revealed through Multi-Omics Studies

Cone snail venom peptides frequently carry a diverse range of PTMs, with up to 75% of amino acids post-translationally modified in a single conotoxin. This heavy and unpredictable incorporation of PTMs often confounds the automated integration of transcriptomic and proteomic data. Current search engines, such as ProteinPilot™ (Sciex, Washington, DC, USA) and the Conomass tool of Conoserver, enable the detection of common conotoxin PTMs including disulfide bonds, C-terminal amidation, pyroglutamylation, glutamate carboxylation, proline hydroxylation, valine hydroxylation, tryptophan bromination, and tyrosine sulfation. However, the identification of complex PTMs like glycosylations is not facilitated by these methods, and their presence usually requires de novo sequencing support. The development of algorithms to facilitate automated PTM identification and verification is needed to facilitate integration, especially for the analysis of big datasets.

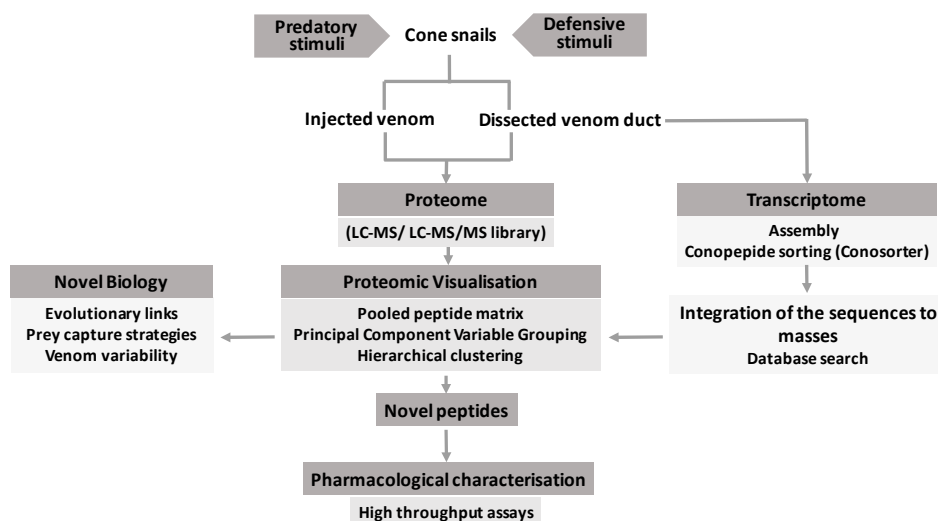
Despite these challenges, completed cone snail venom studies have disclosed an unprecedented level of peptide processing beyond PTMs, including variable peptide processing and transcriptomic messiness. In the first *C. marmoreus* venom study, 7798 unique masses were identified in the proteome, while the venom duct transcriptome revealed only 105 conopeptide coding sequences. Through deep proteomic interrogation, Sebastien et al. uncovered a new mechanism of variable peptide processing (VPP) that contributes to the remarkable diversity of conopeptides [15]. This study highlighted an extensive and highly variable processing of the N- and C-termini of the mature conopeptides that dramatically increased venom peptide diversity. This variable peptide processing, together with PTMs, explains how a limited set of gene transcripts can generate thousands of conopeptides in the venom of a single species of cone snail.

In addition to the variable processing of peptides after translation, a layer of transcriptomic variability was found to arise from low-level transcriptomic variability. Jin et al. found that a surprisingly large number of validated cone peptide gene sequences expressed at low levels contributed to transcriptomic “messiness” and venom hyper-variability [38]. Supporting this variability, Lu et al. found that conotoxin-encoding transcripts can be diversified via hypermutations, fragment insertion/deletions, and mutation-induced premature terminations that allow a single mRNA species to generate multiple toxin products [63]. Thus, both genetic and post-translational messy processing of the venoms contribute significantly to an expansion of conopeptide complexity. All these mechanisms of peptide processing provide a nascent pool of accumulated chemical diversity that can facilitate the rapid evolution of venom peptides with new functions and likely underlies the rapid adaptive radiation found in cone snails.

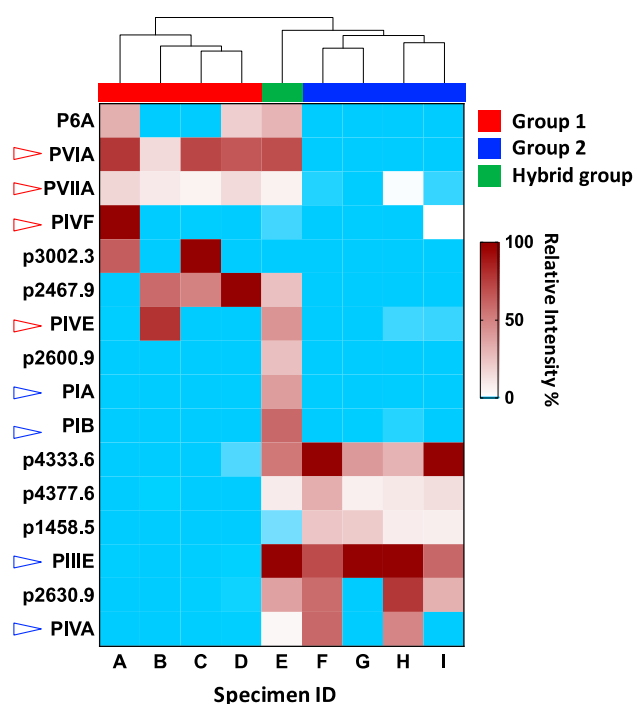
#### *Proteomic Data Visualization Reveals a Remarkable Venom Variability*

Extensive variation in cone snail venom composition is well documented to occur at nearly all biological scales from venom duct regionalisation, to individual variability and species differences. Understanding this variation in venom composition not only would reveal the possible forces that shape venom evolution, but also has implications for venom peptide pharmacology and an understanding of protein structure–function relationships [13]. To better understand this variability, extensive proteomic profiling has been carried out in the past decade [16–21]. The possibility of mass spectra and tandem mass spectra to be readily incorporated into the bioinformatics tools has opened up an exciting opportunity to rapidly profile and visualize this variability to reveal any underlying patterns of peptide expression. Triple-TOF-MS and MS/MS peptide quantification algorithms (Table 3) are increasingly popular for the rapid analysis of complex samples such as venoms. These algorithms, along with clustering techniques such as principal component analysis (PCA) [64,65], hierarchical clustering, molecular networking, and heat-map visualization, provide invaluable tools to gain new insights into venom peptide expression patterns. While it is still early days for data visualization of venoms, this approach has been successfully applied to better understand other complex proteomes as part of a systems biology approach [66–68]. This approach can enrich the integrated venomics approach outlined in Figure 1 to enhance the identification of promising novel peptides that can then be purified or synthesized to discover their functionality.

The first attempt at automating cone snail venom proteomic visualization examined the individual predation-induced venoms from nine *C. purpurascens* [22] using MarkerView™ (Sciex, Washington, DC, USA) to compare and cluster. The downstream analysis with PCA and hierarchical clustering revealed that two major prey-capture cabals (defined sets of synergistic venom peptides) found in this species were typically employed independently and not simultaneously as previously suggested. Interestingly, this clustering approach also revealed separate sets of novel venom peptides associated with either the excitatory or the inhibitory cabals that may help unravel their biology and evolutionary links (Figure 2). While this approach requires better linking to transcriptomic data, it provides a rational and accelerated approach to interrogate the increasing amounts of multi-omics data provided by next-generation proteomics and transcriptomics.



**Figure 1.** A rational venom approach integrating proteomic data visualization approaches to accelerate the comparisons of complex venoms and the rapid identification of likely functionally relevant novel peptides.



**Figure 2.** Proteomic visualisation of the known peptides and most abundant novel peptides of the injected predatory venom collected from nine individual specimens of *C. purpurascens*. The heatmap matrix shows peptides that co-cluster with each of the three groups, revealing remarkable venom variability across these specimens. The red, open arrows indicate known excitatory peptides, and the blue, open arrows indicate known neuromuscular blockers. The relative intensity of each peptide is shown in a gradient of brown, with the absence of peptide indicated by the blue color. This figure is adapted from Himaya et al. 2018.

#### 4. Concluding Remarks

An important goal of cone snail venom research is to accelerate the discovery of structurally diverse and pharmacologically valuable venom peptides and to help unravel their evolutionary trajectories. Already, venomics approaches have identify and understood key evolutionary innovations



of: (i) expanded venom peptide diversity through variable peptide processing and transcriptomic messiness, (ii) regionalization of the venom duct for distinct predatory and defensive functions, and (iii) piscivory and its likely origins from repurposed ancestral worm hunting defensive venoms. However, despite the expansion of sequence information, the rapid and selective pharmacological characterisation of conotoxins and the identification of complex PTMs remain a challenge. To take advantage of the thousands of validated transcriptomic sequences available, a streamlined data-mining, production, and characterization pipeline is required that integrates high-throughput structural and pharmacological characterization with accelerated peptide production systems. Recently, a promising high-throughput recombinant expression of di-sulfide-reticulated venom peptides that can generate large libraries as an alternative to chemical synthesis was reported [69] that may start to fill this growing gap in the pipeline.

**Acknowledgments:** This work was supported by NHMRC Program Grant, NHMRC Senior Principal Research Fellowship (to RJL), and a UQ Postdoctoral Fellowship from The University of Queensland (to S. W. A. Himaya).

**Author Contributions:** S. W. A. Himaya and Richard J. Lewis wrote the paper.

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

## Abbreviations

PTM	Post-translational modification
ACE	Angiotensin-converting enzyme
LC/MS	Liquid chromatography/mass spectrometry
u-HPLC	Ultra-high-performance liquid chromatography
Q-TOF-MS	Quadrupole-time-of-flight mass spectrometry
DDA	Data-dependent acquisition
IDA	Information-Dependent acquisition
VPP	Variable peptide processing
PCA	Principal component analysis

## References

- Harvey, A.L. Natural products in drug discovery. *Drug Discov. Today* **2008**, *13*, 894–901. [[CrossRef](#)] [[PubMed](#)]
- Lewis, R.J.; Dutertre, S.; Vetter, I.; Christie, M.J. Conus venom peptide pharmacology. *Pharmacol. Rev.* **2012**, *64*, 259–298. [[CrossRef](#)] [[PubMed](#)]
- Harvey, A.L. Toxins and drug discovery. *Toxins* **2014**, *92*, 193–200. [[CrossRef](#)] [[PubMed](#)]
- Lewis, R.J.; Garcia, M.L. Therapeutic potential of venom peptides. *Nat. Rev. Drug Discov.* **2003**, *2*, 790–802. [[CrossRef](#)] [[PubMed](#)]
- Pope, J.E.; Deer, T.R. Ziconotide: A clinical update and pharmacologic review. *Expert Opin. Pharmacother.* **2013**, *14*, 9957–9966. [[CrossRef](#)] [[PubMed](#)]
- Nielsen, C.K.; Lewis, R.J.; Alewood, D.; Drinkwater, R.; Palant, E.; Patterson, M.; Yaksh, T.L.; McCumber, D.; Smith, M.T. Anti-allodynic efficacy of the chi-conopeptide, Xen2174, in rats with neuropathic pain. *Pain* **2005**, *118*, 112–124. [[CrossRef](#)] [[PubMed](#)]
- Barton, M.E.; White, H.S.; Wilcox, K.S. The effect of CGX-1007 and CI-1041, novel nmda receptor antagonists, on NMDA receptor-mediated EPSCs. *Epilepsy Res.* **2004**, *59*, 13–24. [[CrossRef](#)] [[PubMed](#)]
- Lubbers, N.L.; Campbell, T.J.; Polakowski, J.S.; Bulaj, G.; Layer, R.T.; Moore, J.; Gross, G.J.; Cox, B.F. Postischemic administration of CGX-1051, a peptide from cone snail venom, reduces infarct size in both rat and dog models of myocardial ischemia and reperfusion. *J. Cardiovasc. Pharmacol.* **2005**, *46*, 141–146. [[CrossRef](#)] [[PubMed](#)]
- Sandall, D.W.; Satkunanathan, N.; Keays, D.A.; Polidano, M.A.; Liping, X.; Pham, V.; Down, J.G.; Khalil, Z.; Livett, B.G.; Gayler, K.R. A novel  $\alpha$ -conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves in vivo. *Biochemistry* **2003**, *42*, 6904–6911. [[CrossRef](#)] [[PubMed](#)]
- Kaas, Q.; Westermann, J.C.; Halai, R.; Wang, C.K.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2007**, *24*, 445–446. [[CrossRef](#)] [[PubMed](#)]

11. 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–D330. [[CrossRef](#)] [[PubMed](#)]
12. Duda, T.F. Differentiation of venoms of predatory marine gastropods: Divergence of orthologous toxin genes of closely related *Conus* species with different dietary specializations. *J. Mol. Evol.* **2008**, *67*, 315–321. [[CrossRef](#)] [[PubMed](#)]
13. Phuong, M.A.; Mahardika, G.N.; Alfaro, M.E. Dietary breadth is positively correlated with venom complexity in cone snails. *BMC Genom.* **2016**, *17*, 401. [[CrossRef](#)] [[PubMed](#)]
14. Dutertre, S.; Jin, A.H.; Vetter, I.; Hamilton, B.; Sunagar, K.; Lavergne, V.; Dutertre, V.; Fry, B.G.; Antunes, A.; Venter, D.J.; et al. Evolution of separate predation-and defence-evoked venoms in carnivorous cone snails. *Nat. Commun.* **2014**, *5*, 3521. [[CrossRef](#)] [[PubMed](#)]
15. Dutertre, S.; Jin, A.H.; Kaas, Q.; Jones, A.; Alewood, P.F.; Lewis, R.J. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteom.* **2013**, *12*, 312–329. [[CrossRef](#)] [[PubMed](#)]
16. Jakubowski, J.A.; Kelley, W.P.; Sweedler, J.V.; Gilly, W.F.; Schulz, J.R. Intraspecific variation of venom injected by fish-hunting *Conus snails*. *J. Exp. Biol.* **2005**, *208*, 2873–2883. [[CrossRef](#)] [[PubMed](#)]
17. Romeo, C.; Di Francesco, L.; Oliverio, M.; Palazzo, P.; Massilia, G.R.; Ascenzi, P.; Polticelli, F.; Schininà, M.E. *Conus ventricosus* venom peptides profiling by HPLC-MS: A new insight in the intraspecific variation. *J. Sep. Sci.* **2008**, *31*, 488–498. [[CrossRef](#)] [[PubMed](#)]
18. Davis, J.; Jones, A.; Lewis, R.J. Remarkable inter-and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* **2009**, *30*, 1222–1227. [[CrossRef](#)] [[PubMed](#)]
19. Dutertre, S.; Biass, D.; Stöcklin, R.; Favreau, P. Dramatic intraspecimen variations within the injected venom of *Conus consors*: An unsuspected contribution to venom diversity. *Toxicon* **2010**, *55*, 1453–1462. [[CrossRef](#)] [[PubMed](#)]
20. Himaya, S.W.A.; Jin, A.H.; Dutertre, S.; Giacomotto, J.; Mohialdeen, H.; Vetter, I.; Alewood, P.F.; Lewis, R.J. Comparative venomics reveals the complex prey capture strategy of the piscivorous cone snail *Conus catus*. *J. Proteome Res.* **2015**, *14*, 4372–4381. [[CrossRef](#)] [[PubMed](#)]
21. Rodriguez, A.M.; Dutertre, S.; Lewis, R.J.; Mari, F. Intraspecific variations in *Conus purpurascens* injected venom using LC/MALDI-TOF-MS and LC-ESI-TripleTOF-MS. *Anal. Bioanal. Chem.* **2015**, *407*, 6105–6116. [[CrossRef](#)] [[PubMed](#)]
22. Himaya, S.W.A.; Mari, F.; Lewis, R.J. Accelerated proteomic visualization of individual predatory venoms of *Conus purpurascens* reveals separately evolved predation-evoked venom cabals. *Sci. Rep.* **2018**, *8*, 330. [[CrossRef](#)] [[PubMed](#)]
23. Dutertre, S.; Lewis, R.J. Cone snail biology, bioprospecting and conservation. In *Snails: Biology, Ecology and Conservation*; Nova Science Publisher's: Hauppauge, NY, USA, 2012; pp. 85–105.
24. Clark, R.J.; Fischer, H.; Dempster, L.; Daly, N.L.; Rosengren, K.J.; Nevin, S.T.; Meunier, F.A.; Adams, D.J.; Craik, D.J. Engineering stable peptide toxins by means of backbone cyclization: Stabilization of the alpha-conotoxin MII. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13767–13772. [[CrossRef](#)] [[PubMed](#)]
25. Clark, R.J.; Jensen, J.; Nevin, S.T.; Callaghan, B.P.; Adams, D.J.; Craik, D.J. The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 6545–6548. [[CrossRef](#)] [[PubMed](#)]
26. Wang, L.; Liu, J.; Pi, C.; Zeng, X.; Zhou, M.; Jiang, X.; Chen, S.; Ren, Z.; Xu, A. Identification of a novel M-superfamily conotoxin with the ability to enhance tetrodotoxin sensitive sodium currents. *Arch. Toxicol.* **2009**, *83*, 925–932. [[CrossRef](#)] [[PubMed](#)]
27. Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212–217. [[CrossRef](#)] [[PubMed](#)]
28. Graves, P.R.; Haystead, T.A. Molecular biologist's guide to proteomics. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 39–63. [[CrossRef](#)] [[PubMed](#)]
29. Jensen, O.N. Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 391–403. [[CrossRef](#)] [[PubMed](#)]
30. Mardis, E.R. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* **2013**, *6*, 287–303. [[CrossRef](#)] [[PubMed](#)]
31. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* **2012**, *131*, 341. [[CrossRef](#)] [[PubMed](#)]

32. Calvete, J.J. Next-generation snake venomomics: Protein-locus resolution through venom proteome decomplexation. *Expert Rev. Proteom.* **2014**, *11*, 315–329. [[CrossRef](#)] [[PubMed](#)]
33. Vetter, I. Development and optimization of FLIPR high throughput calcium assays for ion channels and GPCRs. In *Calcium Signaling*; Springer: Amsterdam, The Netherlands, 2012; pp. 45–82.
34. Calvete, J.J. Venomomics: Integrative venom proteomics and beyond. *Biochem. J.* **2017**, *474*, 611–634. [[CrossRef](#)] [[PubMed](#)]
35. 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. [[CrossRef](#)] [[PubMed](#)]
36. Koua, D.; Brauer, A.; Laht, S.; Kaplinski, L.; Favreau, P.; Remm, M.; Lisacek, F.; Stöcklin, R. ConoDICTor: A tool for prediction of conopeptide superfamilies. *Nucleic Acids Res.* **2012**, *40*, W238–W241. [[CrossRef](#)] [[PubMed](#)]
37. Prashanth, J.R.; Lewis, R.J. An efficient transcriptome analysis pipeline to accelerate venom peptide discovery and characterisation. *Toxicon* **2015**, *107*, 282–289. [[CrossRef](#)] [[PubMed](#)]
38. Jin, A.H.; Dutertre, S.; Kaas, Q.; Lavergne, V.; Kubala, P.; Lewis, R.J.; Alewood, P.F. Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol. Cell. Proteom.* **2013**, *12*, 3824–3833. [[CrossRef](#)] [[PubMed](#)]
39. Aman, J.W.; Imperial, J.S.; Ueberheide, B.; Zhang, M.M.; Aguilar, M.; Taylor, D.; Watkins, M.; Yoshikami, D.; Showers-Corneli, P.; Safavi-Hemami, H.; et al. Insights into the origins of fish hunting in venomous cone snails from studies of *Conus tessulatus*. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5087–5092. [[CrossRef](#)] [[PubMed](#)]
40. Prashanth, J.R.; Dutertre, S.; Jin, A.H.; Lavergne, V.; Hamilton, B.; Cardoso, F.C.; Griffin, J.; Venter, D.J.; Alewood, P.F.; Lewis, R.J. The role of defensive ecological interactions in the evolution of conotoxins. *Mol. Ecol.* **2016**, *25*, 598–615. [[CrossRef](#)] [[PubMed](#)]
41. 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. USA* **2015**, *112*, 1743–1748. [[CrossRef](#)] [[PubMed](#)]
42. Marshall, J.; Kelley, W.P.; Rubakhin, S.S.; Bingham, J.P.; Sweedler, J.V.; Gilly, W.F. Anatomical correlates of venom production in *Conus californicus*. *Biol. Bull.* **2002**, *203*, 27–41. [[CrossRef](#)] [[PubMed](#)]
43. Oldrati, V.; Arrell, M.; Violette, A.; Perret, F.; Sprüngli, X.; Wolfender, J.L.; Stöcklin, R. Advances in venomomics. *Mol. Biosyst.* **2016**, *12*, 3530–3543. [[CrossRef](#)] [[PubMed](#)]
44. Safavi-Hemami, H.; Young, N.D.; Williamson, N.A.; Purcell, A.W. Proteomic interrogation of venom delivery in marine cone snails: Novel insights into the role of the venom bulb. *J. Proteome Res.* **2010**, *9*, 5610–5619. [[CrossRef](#)] [[PubMed](#)]
45. Fox, E.J.; Reid-Bayliss, K.S.; Emond, M.J.; Loeb, L.A. Accuracy of next generation sequencing platforms. *Next Gener. Seq. Appl.* **2014**, *1*. [[CrossRef](#)]
46. Mardis, E.R. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* **2017**, *12*, 213–218. [[CrossRef](#)] [[PubMed](#)]
47. 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. [[CrossRef](#)] [[PubMed](#)]
48. Lluisma, A.O.; Milash, B.A.; Moore, B.; Olivera, B.M.; Bandyopadhyay, P.K. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar. Genom.* **2012**, *5*, 43–51. [[CrossRef](#)] [[PubMed](#)]
49. Lavergne, V.; Harliwong, I.; Jones, A.; Miller, D.; Taft, R.J.; Alewood, P.F. Optimized deep-targeted proteotranscriptomic profiling reveals unexplored *Conus* toxin diversity and novel cysteine frameworks. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3782–E3791. [[CrossRef](#)] [[PubMed](#)]
50. Robinson, S.D.; Safavi-Hemami, H.; Raghuraman, S.; Imperial, J.S.; Papenfuss, A.T.; Teichert, R.W.; Purcell, A.W.; Olivera, B.M.; Norton, R.S. Discovery by proteogenomics and characterization of an RF-amide neuropeptide from cone snail venom. *J. Proteom.* **2015**, *114*, 38–47. [[CrossRef](#)] [[PubMed](#)]
51. Barghi, N.; Concepcion, G.P.; Olivera, B.M.; Lluisma, A.O. Comparison of the venom peptides and their expression in closely related *Conus* species: Insights into adaptive post-speciation evolution of *Conus* exogenomes. *Genome Biol. Evol.* **2015**, *7*, 1797–1814. [[CrossRef](#)] [[PubMed](#)]

52. Decaestecker, T.N.; Vande Castele, S.R.; Wallemacq, P.E.; Van Peteghem, C.H.; Defore, D.L.; Van Bocxlaer, J.F. Information-dependent acquisition-mediated LC–MS/MS screening procedure with semiquantitative potential. *Anal. Chem.* **2004**, *76*, 6365–6373. [[CrossRef](#)] [[PubMed](#)]
53. 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.* **2017**, *129*, 15169–15172. [[CrossRef](#)]
54. Reimers, C.; Lee, C.H.; Kalbacher, H.; Tian, Y.; Hung, C.H.; Schmidt, A.; Prokop, L.; Kaufenstein, S.; Mebs, D.; Chen, C.C.; et al. Identification of a cono-RFamide from the venom of *Conus textile* that targets ASIC3 and enhances muscle pain. *Proc. Natl. Acad. Sci. USA* **2017**. [[CrossRef](#)] [[PubMed](#)]
55. Robinson, S.D.; Li, Q.; Lu, A.; Bandyopadhyay, P.K.; Yandell, M.; Olivera, B.M.; Safavi-Hemami, H. The Venom Repertoire of *Conus gloriamaris* (Chemnitz, 1777), the Glory of the Sea. *Mar. Drugs* **2017**, *15*, 145. [[CrossRef](#)] [[PubMed](#)]
56. Wilkins, M.R.; Sanchez, J.C.; Gooley, A.A.; Appel, R.D.; Humphery-Smith, I.; Hochstrasser, D.F.; Williams, K.L. Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **1995**, *13*, 19–50. [[CrossRef](#)]
57. Mann, M. Origins of mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 678. [[CrossRef](#)] [[PubMed](#)]
58. Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347–355. [[CrossRef](#)] [[PubMed](#)]
59. Zhu, X.; Chen, Y.; Subramanian, R. Comparison of information-dependent acquisition, SWATH, and MSAll techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **2014**, *86*, 1202–1209. [[CrossRef](#)] [[PubMed](#)]
60. Fanali, S. An overview to nano-scale analytical techniques: Nano-liquid chromatography and capillary electrochromatography. *Electrophoresis* **2017**, *38*, 1822–1829. [[CrossRef](#)] [[PubMed](#)]
61. Olsen, J.V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4*, 709–713. [[CrossRef](#)] [[PubMed](#)]
62. Syka, J.E.; Coon, J.J.; Schroeder, M.J.; Shabanowitz, J.; Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528–9533. [[CrossRef](#)] [[PubMed](#)]
63. Lu, A.; Yang, L.; Xu, S.; Wang, C. Various conotoxin diversifications revealed by a venom study of *Conus flavidus*. *Mol. Cell. Proteom.* **2014**, *13*, 105–118. [[CrossRef](#)] [[PubMed](#)]
64. Ivosev, G.; Burton, L.; Bonner, R. Dimensionality reduction and visualization in principal component analysis. *Anal. Chem.* **2008**, *80*, 4933–4944. [[CrossRef](#)] [[PubMed](#)]
65. Ringner, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [[CrossRef](#)] [[PubMed](#)]
66. Meunier, B.; Dumas, E.; Piec, I.; Béchet, D.; Hébraud, M.; Hocquette, J.F. Assessment of hierarchical clustering methodologies for proteomic data mining. *J. Proteome Res.* **2007**, *6*, 358–366. [[CrossRef](#)] [[PubMed](#)]
67. Verhoeckx, K.; Bijlsma, S.; de Groene, E.M.; Witkamp, R.F.; van der Greef, J.; Rodenburg, R.J. A combination of proteomics, principal component analysis and transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line. *Proteomics* **2004**, *4*, 1014–1028. [[CrossRef](#)] [[PubMed](#)]
68. Wiklund, S.; Johansson, E.; Sjöström, L.; Mellerowicz, E.J.; Edlund, U.; Shockcor, J.P.; Gottfries, J.; Moritz, T.; Trygg, J. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **2008**, *80*, 115–122. [[CrossRef](#)] [[PubMed](#)]
69. Turchetto, J.; Sequeira, A.F.; Ramond, L.; Peysson, F.; Brás, J.L.; Saez, N.J.; Duhoo, Y.; Blemont, M.; Guerreiro, C.I.P.D.; Quinton, L.; et al. High-throughput expression of animal venom toxins in *Escherichia coli* to generate a large library of oxidized disulphide-reticulated peptides for drug discovery. *Microb. Cell Fact.* **2017**, *16*. [[CrossRef](#)] [[PubMed](#)]

