

Review

Bioinformatics-Aided Venomics

Quentin Kaas * and David J. Craik

Institute for Molecular Bioscience, the University of Queensland, Brisbane, QLD 4072, Australia;
E-Mail: d.craik@imb.uq.edu.au

* Author to whom correspondence should be addressed; E-Mail: q.kaas@imb.uq.edu.au;
Tel.: +61-7-3346-2021; Fax: +61-7-3346-2101.

Academic Editor: Bryan Grieg Fry

Received: 1 May 2015 / Accepted: 5 June 2015 / Published: 11 June 2015

Abstract: Venomics is a modern approach that combines transcriptomics and proteomics to explore the toxin content of venoms. This review will give an overview of computational approaches that have been created to classify and consolidate venomics data, as well as algorithms that have helped discovery and analysis of toxin nucleic acid and protein sequences, toxin three-dimensional structures and toxin functions. Bioinformatics is used to tackle specific challenges associated with the identification and annotations of toxins. Recognizing toxin transcript sequences among second generation sequencing data cannot rely only on basic sequence similarity because toxins are highly divergent. Mass spectrometry sequencing of mature toxins is challenging because toxins can display a large number of post-translational modifications. Identifying the mature toxin region in toxin precursor sequences requires the prediction of the cleavage sites of proprotein convertases, most of which are unknown or not well characterized. Tracing the evolutionary relationships between toxins should consider specific mechanisms of rapid evolution as well as interactions between predatory animals and prey. Rapidly determining the activity of toxins is the main bottleneck in venomics discovery, but some recent bioinformatics and molecular modeling approaches give hope that accurate predictions of toxin specificity could be made in the near future.

Keywords: toxins; databases; algorithms; proteomics; transcriptomics; phylogeny; molecular modeling

1. Introduction

Animals that use venoms for killing or paralyzing prey or for defense purposes are widespread in diverse phyla of the animal kingdom. They are found in nearly all major animal subgroups, including vertebrates (e.g., snakes, lizards, fish, toads, frogs, vampire bats, platypus, echidna), mollusks (e.g., marine cone snails, terebrids, cuttlefish, octopus, squids), arthropods (e.g., scorpions, spiders, centipedes, ticks, ants, wasps, bees), annelids (e.g., leeches), sponges and cnidarian (e.g., jellyfish, sea anemones). Animal venoms are a complex mixture of substances, mainly composed of proteins and peptides, but also containing lipids, amines and other small molecules [1]. Venom peptides have sparked particular interest because of their activity on the nervous system (blockers or agonists of ion channels, transporters or receptors) or on membranes (antibacterial, hemolytic). Larger venom proteins have enzymatic activity (phospholipase A2, proteases, and oxidases) or modify important physiological process such as coagulation (lectins). The venom of a single species can contain hundreds to several thousands of active peptides [2,3], and the total pool of unique active peptides in venoms is therefore extremely large, probably greater than ten million. For example, hundreds of thousands of different active peptides have been proposed to exist in the venoms of the greater than 700 marine cone snail species [3–5]. Other species extensively studied for their venom content belong to more highly diversified orders or sub-orders; for example, there are ~45,400 species of spiders [6], ~3500 species of snakes [7,8] and ~2000 species of scorpions [9].

Animal venoms are one of the richest natural sources of active compounds and have numerous pharmacological and medical applications [10]. Toxins are used as pharmaceutical tools (e.g., to characterize ion channels), and also in pharmaceutical applications (e.g., drugs) or biotechnological applications (e.g., insecticides) [10]. The cone snail toxin MVIIA is an analgesic drug targeting voltage-gated calcium channels (Cav), which is used under the name Prialt[®], to treat intractable pain [11]. Other toxins are also in development as treatments of various nervous system disorders (epilepsy, pain), as well as arthritis, heart and blood pressure disorders and cancer [12,13]. Considerable effort is undertaken to discover new toxins with interesting activity, and in the course of this research, extensive knowledge on how toxins are produced has been gained.

In most studied organisms, peptide toxins are stored in venom glands. The cells lining these glands produce toxins as precursor proteins, which undergo several maturation steps before being secreted [14]. Toxin precursors are first directed to the endoplasmic reticulum (ER), where the ER signal peptide region is cleaved. While transiting through the secretory pathway, the truncated toxin precursors mature by undergoing several post-translational modifications (PTMs), which include excision of the mature region, formation of disulfide bonds between pairs of cysteine residues and modifications of side chains or peptide termini, for example hydroxylation of prolines or amidation of the C-terminus [15–17]. The formation of disulfide bonds is a particularly important PTM because it stabilizes the three-dimensional (3D) structures of toxins, and consequently the number and arrangement of the disulfide bonds is commonly used to classify toxins [15]. (The many toxins that do not have disulfide bonds are typically classified in families according to their sequence similarities). Overall, 15 side-chain PTMs have so far been identified in cone snail toxins [15]. These modifications contribute to increasing the chemical diversity of venoms, fine-tuning the specificity for molecular targets or helping the maturation of toxins [18]. Studies of toxin evolution have revealed that most

animal toxin genes belong to gene superfamilies that probably emerged from duplication of a few ancestral genes [19]. The ER signal peptide of toxin precursors shares high sequence similarity within a gene superfamily but the mature peptide regions undergo rapid evolution and are consequently more divergent [16]. This fundamental knowledge of toxin maturation and evolution has led to the development of targeted approaches to discover them.

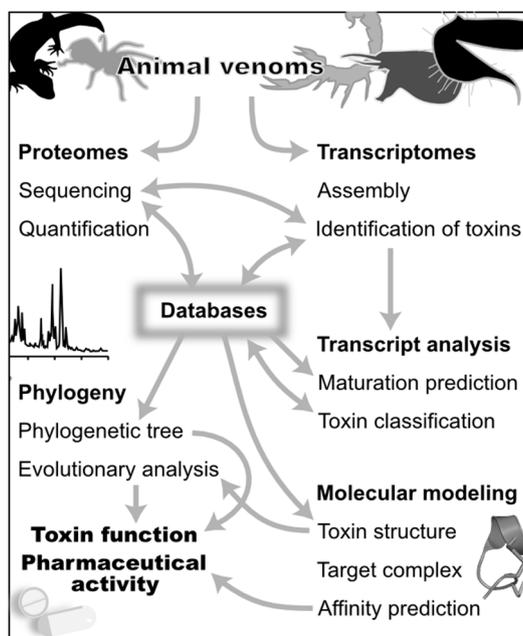


Figure 1. Computational approaches in venomomics. The different themes in this figure are discussed in the text. Generalist and specialized databases are a central source of information for both the discovery (**top**) and analysis (**bottom**) of venoms. Venomics computational tools combine data from proteomics and transcriptomics to discover the set of toxins in a venom gland. These tools use predictions of mature toxins from transcriptomes to support peptide sequencing, and sequencing by mass spectrometry uses transcript sequences as a database to rapidly identify peptides and their post-translational modifications. Bioinformatics tools use the standardized classification stored in databases to analyze transcripts, and phylogenetic analyses can be used to analyze toxin evolutionary relationships. Databases in turn record the newly identified toxin peptide and transcript sequences. These sequence data can be also used to refine toxin classification, for example when a new phylogenetic group of toxins is identified. The molecular target of toxins can be suggested by a phylogeny and an evolutionary analysis complemented by the prediction of toxin 3D structures. Molecular modeling of target/toxin complexes can be used to analyze in great depth structure-activity relationships of toxins. Finally, the affinity of these complexes can be predicted from the molecular models.

Over the last few years, cutting-edge technologies are increasingly being employed in toxin research to unravel the pharmacological treasures hidden in animal venoms, dramatically changing the landscape of venom exploration [10]. Combined advances in proteomics and transcriptomics now give access to nearly complete toxin repertoires of a single venom [20], with this new approach termed “venomics” [2,10,21]. Venomics research is driven by techniques rather than by hypotheses [10], with

old fractionation methods now replaced with the simultaneous discovery of massive amounts of toxin nucleic and protein sequences. This large amount of data cannot be efficiently comprehended without the use of computational and statistical methods [22]. This review will first focus on computational approaches that have been developed in recent years to help in the discovery and analysis of venom proteins and peptides. The main themes discussed are illustrated in Figure 1. Several specialized databases provide access to information on toxins, which are used by bioinformatics tools to analyze “omics” data and to orient new venomomics research. The increased pace at which toxin sequences are being discovered is unfortunately not matched by the speed at which their functions are assayed because the functional characterization of a single toxin requires significant experimental efforts. Thus, we also describe some of the computational approaches that have been developed to suggest and study toxin functions.

2. Databases

Sequence and structure data on toxins are dispersed and difficult to extract from generalist sequence databases such as GenBank [23] because toxin sequences are not always annotated as such or because sequences are highly redundant [22]. Generalist databases rely on authors to submit their sequences, and consequently a large amount of information is in the peer-reviewed literature but not in generalist databases. Some aspects of toxin-centered databases were reviewed recently [24]. ConoServer (<http://www.conoserver.org>), ArachnoServer (<http://www.arachnoserver.org>), and ISOB (Indigenous Snake species Of Bangladesh, <http://www.snakebd.com/>) provide expert annotations on the sequences and 3D structures of cone snail, spider, and snake toxins, respectively [25–27]. There are currently no equivalent databases for other venomous species, and this absence has been noted as impeding venomomics efforts [1]. ConoServer, ArachnoServer and ISOB catalog data that are either manually extracted from the peer-reviewed literature or semi-automatically retrieved from generalist sequence and 3D structure repositories such as UniProtKB [28], GenBank [23], GenPept and the Protein Data Bank (PDB) [29]. The origin of each entry in ConoServer, ArachnoServer and ISOB can be traced back to the peer-reviewed literature and/or to a generalist database. ConoServer and ArachnoServer databases solve naming issues related to toxins that have several names in the literature, ultimately giving names to toxins according to cone snail and spider toxin standard nomenclatures [15,30]. Another important task of these two databases is to keep track of the rapidly evolving toxin classifications, which are frequently amended. In the remainder of this section we now give more information on the five most useful publicly available sources of data in venomomics, namely ConoServer [25,31], ArachnoServer [26,32], ISOB [27], Tox-Prot [33] and ATDB (Animal Toxin DataBase) [34].

ConoServer [25,31] catalogs 2201 mature toxin sequences, 2837 precursor sequences and 175 3D structures of cone snail toxins (as of June 2015), which have been retrieved from public repositories and from the peer-reviewed literature. Mature peptide toxin entries, protein and nucleic acid precursor entries, and 3D structure entries are linked. All mature peptide toxins, discovered experimentally or predicted from the precursor, are entered as independent entries. The prediction of mature peptide toxins allows identifying precursor sequences that mature into the same toxin, discovering these peptides by mass-spectrometry (MS), and aligning the mature peptides on the database website.

The experimental evidence supporting the existence of each toxin, *i.e.*, if it was discovered at the protein or nucleotide level, is provided. ConoServer only records experimentally identified PTMs, and these are graphically annotated on the sequences of the mature toxins. The biological activity on molecular targets are described in terms of IC₅₀, K_i, and percentage of inhibition, and the targets are named according to the classification of the International Union of basic and clinical PHARmacology (IUPHAR) [35]. The function of a toxin can be queried using the “pharmacological family” classification, which describes the type of molecular target and the type of activity, for example the κ -conotoxins describe cone snail toxins that block voltage-gated potassium channels. NMR models are retrieved from the Biological Magnetic Resonance Bank (BMRB) [36] but some are also directly submitted to ConoServer by authors upon request. The database also keeps track of patented sequences as well as information on synthetic peptides derived from cone snail toxins; for example, synthesized for structure-activity relationship studies. Statistics on the classification schemes, PTMs, and toxins per cone snail species are automatically updated. Each entry has a stable identifier, which comprises the letter “P” for proteins or “N” for nucleic acids followed by a number. The peer-reviewed references that were used to create the database are listed on a single page, providing a convenient bibliography on cone snail toxin discovery. Database sequences can be retrieved in FASTA format, and all annotations can be downloaded in XML format. The database is regularly updated.

ArachnoServer [26] catalogs 1464 mature spider toxins, which were retrieved from UniProtKB [28] (as of June 2015). Mature toxin PTMs are provided and clearly identified as being experimentally validated or predicted by homology. Disulfide bond connectivities are difficult to determine experimentally, and ArachnoServer supplies an experimental confidence level for them. The toxin molecular targets are described according to the IUPHAR nomenclature [35], and the toxins can be browsed according to their molecular targets using a powerful interactive tool. ArachnoServer provides experimental values of target inhibition for a large number of entries, and considerable efforts have been made to provide comments on toxin functional characterization. Enzymes are described using the Enzyme Commission (EC) number and are linked to the MEROPS database [37]. All mature toxins in ArachnoServer are named according to the standard nomenclature for spider toxins [30], but alternative names found in the literature are also recorded. Each toxin entry has a unique identifier, which comprises the two letters “AS” followed by a number. Sequences can be retrieved in FASTA format, and every list of entries can be saved in XLS or PDF formats.

ISOB (<http://www.snakebd.com/>) [27] catalogs toxins from snakes in Bangladesh. This recently created database can be queried by the snake species scientific name as well as by toxins. The database interface does not allow searching toxins by name or sequence but they can be listed according to their activity, such as serine proteinase, neurotoxin or phospholipase. Information provided for each toxin includes its name, functional class, uniprot accession number, reference, protein sequence and structural information. When no experimental structure is available, a homology model is generated for each toxin. As of June 2015, 76 species of snakes and 419 proteins and toxins are listed in ISOB.

Tox-Prot (<http://www.uniprot.org/program/toxins>) [33] is the animal toxin annotation program of UniProtKB and aims to systematically annotate animal venom peptides [24]. Some of the annotations found in ConoServer and ArachnoServer have been transferred by Tox-Prot into UniProtKB, and links between the databases have been created. There are fewer annotated entries in UniProtKB concerning cone snail toxins than in ConoServer; for example only 1011 cone snail toxin sequences are annotated

in UniProtKB. Each toxin entry in UniProtKB displays less annotation on the functional characterization than ArachnoServer or ConoServer. Besides cone snail and spider toxins, Tox-Prot stands as the best choice to retrieve information on toxins for the majority of other venomous species. Toxin activities are described in comment fields in UniProtKB entry cards. Two ontologies are used to describe ion channel and receptor subtypes: IUPHAR [35] and Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) ontologies. Two other ontologies describe enzymes found in venoms: EC numbers and the nomenclature from the International Union of Biochemistry and Molecular Biology (IUBMB).

The animal toxin database (ATDB, <http://protchem.hunnu.edu.cn/toxin/>) [34] is a meta-database that displays information from several databases, including UniProtKB, GenPept and three discontinued databases dedicated to scorpion toxins (SCORPION2; [38]), cone snail toxins (MOLLUSK) and some snake toxins (svPLA2). ATDB is based on highly detailed ontologies describing the toxins and the target ion channels. ATDB toxin ontology describes the function of the toxins at different levels: The mode of actions and effects on the molecular targets, the symptoms, the pathogenesis (e.g., “Cytolysis/Cell degranulating/Muscle necrosis”) and the functional type of the toxins (e.g., “Neurotoxin/Postsynaptic neurotoxin”). The ion channel ontology describes the targeted ion channels according to the IUPHAR classification [35], distribution, physiology (e.g., “Fast deactivation”) and pharmacology (e.g., “System effects/Pain responses/Inflammatory pain”). These detailed ontologies would potentially allow very clever browsing of the database, but the interface currently fails to display the corresponding entries for the ion channel ontology. The database was last updated in 2009.

3. Discovery of Toxins from Venom Gland Transcriptomes

Because of the rapidly decreasing cost of sequencing nucleic acids, especially using massively parallel sequencing technologies, a large number of transcriptomic analyses of snake, cone snail, spider, scorpion and a few other animal venom glands have been carried out. Besides the discovery of pharmaceutically active compounds, these studies are also motivated by academic interests in the genetic mechanisms at the origin of toxin families [19] or in the co-evolutions of venoms and preys [39,40].

Only a few studies have reported the sequencing of the genomes of venomous species, which is a major undertaking because eukaryote genomes are typically large and have numerous tandem repeats. Currently, two snake genomes (Burmese python and king cobra) [41,42], two cone snail genomes (*Conus bullatus* and *C. consors*) [43,44], one scorpion genome [9], one spider genome [45], and the genomes of the honeybee [46] and parasitic wasps [47] were reported as being at various stages of completion. By contrast, a large number of studies have reported sequencing of the venom gland transcriptomes, which are more affordable and the bioinformatics analyses are easier than for sequencing genomes. Consequently, we focus here mainly on the analysis of transcriptomes rather than genomes.

The analysis of the transcriptome of a venom gland proceeds in several steps, which differ only slightly from the standard procedure for other tissues [48]: The sequencing reads are first assembled into contigs, the functions of contig gene products are predicted by homology, toxin precursors are identified among the contigs, toxin precursors are analyzed and classified, and reads and contigs are submitted to generalist repositories, such as NCBI SRA, dbEST and/or GenBank [49].

From a bioinformatics point of view, the important characteristics of the different sequencing technologies are the length of the reads compared to those of the transcripts, the number of reads, and the sequencing error rate. Second generation sequencing technologies, such as 454 GS FLX Titanium (454) and Illumina, give access to the sequences of lowly expressed transcripts that were undetected by Sanger sequencing methods [50], but the new technologies have a much higher rate of sequencing errors. For example, the 454 has an average error rate of 1% [51], and studies that claim observing transcript variations supported by only one sequencing read [52], should be viewed with caution. Newly discovered transcripts should ideally be supported by proteomics data or other nucleic acid sequencing technologies [50].

Sequencing reads are typically assembled *de novo* into larger contigs, and this procedure requires discarding low-quality reads. Illumina reads are typically first analyzed for their quality using the Illumina quality scores, and this analysis orients the choice of the quality threshold of reads as well as the trimming of low quality bases at the 3' end of all reads. ESTs can be assembled into contigs using various software, the most popular being CAP3 [53], Phrap [54] and SeqMan [55]. The most popular assemblers for 454 reads are MIRA [56], Newbler (Life Science, Frederick, CO, USA) and CLC Genomics Workbench (CLC Bio, Aarhus, Denmark), and popular transcriptome assemblers used with Illumina reads are Velvet/Oases [57] and Trinity [58]. *De novo* assembly, *i.e.*, without a reference genome, of second-generation sequencing reads is still considered challenging and should be treated with caution. Multigene toxin families display high similarity in part of their sequences [16], and are therefore particularly difficult to assemble because similar or even identical read sequences and a certain rate of sequencing error are potentially shared between several transcripts. Issues associated with second-generation sequencing assemblies are the failure to assemble certain transcripts and the creation of spurious contigs (chimeric transcripts). Some studies assess the reliability of assemblies by cross validating results from several *de novo* assemblers [59], but such strategy is highly time consuming and not totally conclusive. The choice was made in some studies to avoid assembly problems and to search for toxin transcripts directly in the sequencing reads. The distribution of open reading frames of cone snail toxin transcripts is indeed similar to the distribution of read lengths produced by 454 sequencing technology [50,60]. VTBuilder is a recently developed transcriptome assembler that considers relationships between co-evolving sites to avoid building chimeric transcripts, and has been shown to be effective at assembling toxin transcripts from Illumina reads [61].

Contigs corresponding to toxin transcripts can be identified by similarity with known toxin sequences. The potential function of non-toxin contigs is predicted by homology with entries in NCBI/nr [49] and/or UniProtKB [28]. Standard annotations of non-toxin sequences include the enrichment in gene ontology terms using Blast2GO [62], and the identification of protein families, domains and important sites using InterProScan [63]. Proteins belonging to certain metabolic and regulatory pathways can also be identified using the KEGG annotation server [64] and iPath explorer [65]. This approach was recently used to suggest that certain metabolic pathways are not directly involved with toxin production. For example, one study found that fatty acid synthesis is down-regulated in the venom gland of the cone snail *Conus geographus* [66].

Toxin precursors belonging to different gene superfamilies are highly divergent, and toxins belonging to previously undescribed gene superfamilies are difficult to identify using sequence similarity searches. Profile-based method alignments are more sensitive and have been used in a few

studies analyzing venom gland transcriptomes [52,67,68]. For example, Shwartz *et al.* [68] used a sequence profile built with PSI-BLAST [69] to suggest the remote homology between toxin venom gland transcripts and other scorpion toxins that are potassium channel blockers. Profile-hidden Markov models (pHMMS) were recently used to identify toxin transcripts in cone snail transcriptomes [52,67,70,71]. ConoDictor uses pHMMs as well as generalized position-scoring matrices to classify cone snail toxin precursors into gene families or superfamilies [70,71], and this method has been used to identify toxins from transcriptomic data [72]. The sequence profiles of ConoDictor were constructed for signal peptide, pro-region and mature peptide region sorted into different groups based on sequence similarity. ConoDictor is an online tool (<http://conco.ebc.ee>), and it is therefore convenient to use. ConoSorter [52] uses several pHMMs to rapidly identify toxin transcripts in 454 reads or assembled contigs. Similarly to ConoDictor, the pHMMs of ConoSorter were created using HMMER [73], which was run on alignments of manually created clusters of transcript regions (signal peptide, pro-region and mature peptide region) for each gene superfamily. Alignment clusters were created manually because conventional clustering tools failed to produce reliable results, highlighting the extreme divergence between these sequences [52]. ConoSorter is a computer script, which needs to be run locally and can handle very large number of sequences. A simpler approach was recently employed by Robinson *et al.* [67]: One pHMM was created for each gene superfamily using alignments of all non-redundant sequences in ConoServer, without creating clusters [25]. Other approaches to identify toxins in assemblies or reads are based on the knowledge that toxins display a limited number of cysteine patterns in their amino acid sequence [59] or adopt a limited number of folds [74]. For example, giant ant *Dinoponera quadriceps* toxins were identified by local sequence alignment against the KNOTTIN database [75], which catalogs proteins adopting a cystine knot fold [76]. Knoter 1D is an on-line tool that automates this process by predicting cystine knot folds based on sequence similarity to known knotted proteins as well as a range of other features, such as the number of cysteine residues, the length of sequences between consecutive cysteine residues or annotation keywords [75]. Besides sharing similar folds, toxins acting at similar molecular targets should display similar sequence features, and ClanTox [77] is a machine learning approach that aims to identify ion channel inhibitors using sequence features only. The methods implemented in ClanTox combine the results of several decision-stump classifiers using the AdaBoost algorithm [78] (boosted-stump classifiers). ClanTox is species independent, as opposed to ConoDictor and ConoSorter, and can be used to identify toxin-like sequences. For example, ClanTox was used to discover toxin-like proteins in mammalian transcriptomes [79], and some of the identified peptides are modulators of cell signaling acting at nicotinic acetylcholine receptors using a three-finger snake toxin fold [80,81]. On the basis of convergent evolution of toxins in different phyla, Starcevic *et al.* identified sequences motifs that seem to be specifically used by toxins [82]. They used MEME [83] to identify 748 shared motifs of about 40 amino acids in length, and the pHMM built from alignment of these motifs was shown to be able to identify snake PLA2 toxins among other proteins and to generate only a limited number of false positives when analyzing transcriptomes of non-venomous animals.

4. Discovery of Toxins from Proteomes

Most of the known toxin peptide sequences were predicted from conceptual translation of nucleic acid sequences. Consequently, the majority of toxins cataloged in databases do not have any experimental support for them being produced in venoms. For example, only 379 out of the 1873 mature toxins recorded in ConoServer have some experimental evidence of being present in venom. Nevertheless, support for mature peptide toxin sequences becomes more routinely and rapidly provided by modern proteomics experiments, e.g., [20,60]. A major issue in toxin proteomics derives from large intra-species and even intra-specimen variability, indicating that understanding the biology of the venomous animals is essential to characterize venom content [84,85]. Additionally, each venom displays a very large chemical diversity, which can be explained by recently evidenced mechanisms, including differential post-translational modification of toxins [86], differential truncations [60], and transcription messiness [50]. Transcription messiness is an emerging concept where errors of transcriptions cause epimutations, resulting in several proteins with slightly different amino acid sequences [87–89].

Proteomics approaches to analyze snake venom were recently reviewed [1,90], and the techniques discussed in these reviews are generally applicable to other animal venoms. Venoms typically contain hundreds of different peptide components [3], and a wide range of combination of chromatography, electrophoresis, enzymatic digestions, Edman degradation, or mass spectrometry (MS) techniques have been employed to sequence their toxins [1,90]. Edman degradation is the traditional approach to sequence peptides, and it still sometimes used for full [91] or partial [92] toxin sequencing. Most toxins are very small and can be sequenced at lower cost using tandem MS (MS/MS) [50,60,90], but Edman degradation can be useful as a complement to MS, e.g., reference [93], for example helping to identify the isobaric amino acids isoleucine/leucine and for *N*-terminal sequencing.

Peptide sequences are typically discovered from MS/MS data by (i) complete *de novo* sequence reconstruction, (ii) reconstruction of some partial sequence tags to be searched in a sequence database, or (iii) matching spectra predicted from a database of sequences. Complete *de novo* sequencing is slow and its challenging nature is exacerbated by the high frequency of side-chain PTMs. Despite the creation of some specific methodologies [94], this strategy is no longer used routinely. The identification of peptides in MS/MS spectra using proteins predicted from transcriptomes is now standard procedure [50,60,95]. This proteogenomic strategy leads to a higher level of confidence for the mature peptide toxin sequences because transcriptomes provide independent validation and also help to resolve ambiguities regarding the identification of isobaric residues, *i.e.*, isoleucine and leucine residues. The majority of side chain PTMs cannot be predicted from transcript sequences, and these PTMs potentially render the identification of peptides more difficult, but the widely used software for matching MS data with sequence databases, Mascot (Matrix Science, Boston, MA, USA), ProteinPilot (Applied Biosystems, Waltham, MA USA), SEQUEST (Thermo Finnigan, San Francisco, CA, USA), and x!tandem [96] readily take into account a very large number of PTMs. The latest developments in the field have been recently reviewed [97,98], and we will focus on bioinformatics methods specifically made to analyze venom content. A tool called ConoMass, accessible from the ConoServer website, uses a brute force approach to match reconstructed peptide masses from MS data with toxin masses predicted from transcriptome analysis [25,60]. ConoMass considers all possible combinations

of PTMs that were identified in cone snail toxins to date, with the exception of side chain glycosylations, which are very diverse in their occurrence [25,99]. We note that alternative disulfide bond connectivities, which can lead to variation in bioactivity [100], cannot be detected via mass changes. The identification of peptide using such a simple method, which is also employed by other algorithms such as FindMod [101], typically generates a large number of false positives [102]. Increased sensitivity of mass spectrometers as well as knowledge of PTMs commonly found in related peptides from the same gene superfamily could help to discriminate between alternative toxin sequences for each reconstructed mass [60].

The quantification of toxins in the venom can be carried out by integration of the area under chromatography elution curves. Interestingly, a recent comparison of the quantification from proteomes and transcriptomes in the venom gland of a snake showed large discrepancies, possibly because once the concentration of a toxin in the glands reaches a certain level, its production is repressed [103]. Some toxins predicted to exist from transcriptome data are not found in the proteome, and conversely toxins in the proteome have no identified transcripts. It is therefore important to cautiously interpret the expression level of transcripts discovered from a transcriptome analysis in terms of importance in the venom. Possible explanations of these discrepancies have been proposed [1]: Some toxins might not be secreted and will not be found in the venom; the origin of the material is different (several venom samples are typically pooled from several animals but RNA material is only collected from one specimen); the sensitivity of proteomics is still too low; and some types of side chain PTMs might still be unknown. An alternative explanation was proposed in a recent study of the origin of the ontogenic shift in venom content of the Central American rattlesnake [104]. miRNA levels were shown to be the main factor that modulates venom composition as the relative toxin transcriptional activity was similar whatever the development stage.

5. Bioinformatics Analysis of Toxin Transcripts and Classifications

The analysis of toxin transcripts typically proceeds by determining the different regions in the precursors (ER signal peptide, mature peptide region and pro-regions) as well as the estimation of their evolutionary relationships and the structural and functional class of their mature toxins. With the increasing number of known toxins, the majority of these tasks can be done by transferring annotations by homology. The ER signal peptide boundary is often predicted using SignalP [105], which is currently the best method to predict the presence of an ER signal peptide (Mathew correlation coefficient of 87%) but is less accurate at predicting the exact position of the cleavage (Mathew correlation coefficient of 68%) [105]. Using the knowledge that the ER signal peptide is highly conserved within gene superfamilies, an alignment of sequences can alternatively be used to predict the ER signal peptide cleavage site. The determination of the mature peptide region in the precursor sequence is far more complicated, because we still have limited knowledge on most proprotein convertases specific cleavage sites [106]. A “rule of thumb” that has been used in some studies is to assume that a cleavage after all basic residues, *i.e.*, arginine or lysine residues, surrounding the sequence region containing all cysteine residues, *e.g.*, [107]. This simple strategy captures very approximately the cleavage sites of furin-like peptidases, leading to about 50% wrong predictions for Cone snail precursors of cysteine rich toxins, and cannot be used for toxins that do not have cysteine

residues. A more rigorous approach was implemented in ProP, which is an artificial neural network trained to recognize proprotein convertase cleavage sites [108]. ProP is highly accurate at recognizing furin-like protease cleavage sites, with 89% balanced accuracy on a large dataset, but has only an average 61% balanced accuracy when considering all types of proprotein convertases. Only 113 out of 291, *i.e.*, 39%, of experimentally validated cleavages sites surrounding mature cone snail toxin regions correspond to furin-like protease cleavage sites, and therefore using ProP to annotate cone snail toxin transcripts leads to a significant number of wrong predictions. The number of furin-like cleavage sites in spider toxin transcripts is very low, and ProP achieved a sensitivity of 0% in a recent study focusing on spider toxins [109]. Two specific tools, ConoPrec [25] and SpiderP [109], have been developed to annotate cone snail and spider toxins, respectively.

ConoPrec is available on the ConoServer website. Cone snail toxin precursor sequences can be provided to ConoPrec either as nucleic acid or protein sequences. This tool first determines ER signal peptide region using SignalP and then mature peptide regions using a set of regular expressions corresponding to proprotein convertase cleavage sites [106]. These regular expressions match cleavage sites of the well-studied furin proteases as well as cleavage sites from other protease families, the specificity of which were only suggested from statistical analysis of alignments of precursor sequences [106]. The protease at the origin of some of these cleavages have not yet been identified, such as for the “EtoR” cleavages, which do not exist in mammals [106]. Some cleavages seem to be specific of some species, for example the “LtoR” cleavages for which a cone snail proprotein convertase was isolated [110]. ConoPrec achieved ~80% balanced accuracy for the identification of the 291 verified cone snail toxin cleavage sites in ConoServer, and it was also recently benchmarked to have similar balanced accuracy for analyzing spider toxin precursors [109]. The ability to achieve high accuracy in analyzing precursors from animals other than cone snails indicates that ConoPrec could potentially be used generally for all animal venoms. A current limitation of ConoPrec is that it only predicts cleavage sites of disulfide-rich toxin precursors.

SpiderP is a recently developed tool that uses a support vector machine (SVM) to predict the location of proprotein convertase cleavage sites in spider toxin precursors [109]. The SVM of SpiderP uses Gaussian Radial Basis Function (RBF) kernels and was trained to recognize cleavage sites on an eight residue long window. A negative dataset of 8-mers was created using spider sequences that are not proprotein convertases. On an independent dataset of 46 spider toxin protein precursors, SpiderP achieved an impressive balanced accuracy of 90%. The absence of furin-like cleavage sites in spider precursors indicates that this tool will be less efficient at analyzing transcriptomes in animals other than spiders. The algorithm can be used on-line on the ArachnoServer website.

Cone snail toxins are classified (i) in terms of gene superfamilies; (ii) according to the pattern of their cysteine residues in the primary sequence, which partly reflects the 3D fold of the mature toxins; and (iii) using their pharmacological activity, often named with a Greek letter, *e.g.*, α -toxins, are blockers of nicotinic acetylcholine receptors. Classifying newly discovered toxins is an important task, and ConoPrec classifies cone snail toxin precursors into cysteine frameworks using their pattern of cysteines, and it assigns precursors to a gene superfamily if their ER signal peptide is more than 90% identical to that of any member of known cone snail gene superfamilies [25]. ConoPrec also determines the closest cone snail toxin in the ConoServer database, and this homolog toxin could be

used to suggest a pharmacological activity. Three PTMs can be predicted on the basis of specific sequence features, including pyroglutamylation of *N*-terminal glutamine, *C*-terminal amidation and γ -carboxylation of glutamates.

The conserved ER signal peptide sequence provides definitive support for classifying toxin transcripts into gene superfamilies. Bioinformatics tools have been developed to classify cone snail toxins into gene superfamilies from partial sequences missing the ER signal peptide. ConoDictor and ConoSorter [52,70,71], which were introduced in the section describing the identification of toxins from transcriptomic data, determine the closest toxin gene superfamilies using pHMMs and/or generalized position specific scoring matrices. [71] Other studies have applied machine-learning approaches to extract the mature toxin sequence patterns associated with each cone snail gene superfamily. For example, a strategy using pseudo-amino acid composition and multi-class SVM achieved 88% correct predictions [111]. The method called IDQD uses an algorithm based on Mahalanobis discriminant instead of a SVM and achieved a balanced accuracy of 90% as well [112]. Another study used diffusion maps to optimize prediction features, and achieved an overall accuracy of 90% [113]. A recent study used an original approach by considering the similarities between small subsequences of each sequence compared to the total pool subsequences [114]. This study used SVM as a classifier and achieved a balanced accuracy of 98%, according to the authors.

6. Phylogenies and Evolutionary Analysis

Transcriptome analyses are often complemented by the establishment of molecular phylogenies between related toxin transcripts. Because structure and function tend to be conserved through evolution, related toxins display similar properties and a gradation of traits is expected to be associated with increasing phylogenetic distances between toxins [115]. For example, snake toxins displaying a three-finger scaffold have taxon-specific activity, and toxins belonging to distantly related branches display different receptor subtype or target specificity [116]. Establishing the phylogeny of toxin sequences could therefore help in determining their function. Describing the evolution of venoms relative to prey has also been used to inform the discovery of new toxins with interesting pharmaceutical activity [117]; for example, by focusing on studying the venom of fish-hunting cone snails [118], or on clades whose venom pool remained so far unexplored [119].

Studying the evolution of toxin scaffolds brings some understanding on how folds evolved relative to their function. A recent study of scorpion toxins proposed that the Von Willebrand factor *C*-domain fold (SV-SVC) is the ancestral fold from which the inhibitory cystine knot fold (ICK) evolved, which in turn gave rise to the disulfide-directed β -hairpin fold (DDH) [39]. Scorpion ICKs and DDHs have both been extensively studied for their function, and DDH toxins are more potent on human ryanodine receptors than ICK toxins, indicating that selection pressure results in more active compounds [39]. The loss of cysteine residues 2 and 3 in a clade of snake three-finger toxins led to a greater potency on mammalian receptors, which correlated with change of diet and explosion of new toxins displaying this modification [116,120]. Hence, molecular phylogeny could help pharmacological investigations of animal toxins by suggesting the clades the most likely to contain pharmaceutically interesting compounds.

6.1. Building Phylogenetic Trees

Classical phylogenetic analyses of protein or nucleic sequences require the generation of a multiple sequence alignment, the accuracy of which has a critical influence on the resulting phylogeny. Toxin sequences are often substantially divergent and generating a reliable multiple sequence alignment is difficult. Some authors use sequence alignments generated by software such as ClustalW [121], which was used to establish the phylogeny of some spider toxins [122] or Muscle [123], which was used to study the evolutionary relationships between some cone snail toxins [124]. Other authors rely on manual alignments [125]. A common strategy to reduce noise due to regions with unreliable alignment is to delete them from the multiple alignments. In some studies, all position displaying with gaps are deleted [122], and in some others only positions above a percentage of gaps are removed [39,116]. In the study of Sunagar *et al.* [39], two alignments of ICK and DDH sequences, corresponding to two different evolutionary scenarios, were considered, and the main conclusions from the two resulting but slightly different phylogenies globally agreed.

The methods most frequently used to build phylogenetic trees are the neighbor-joining algorithm, maximum likelihood, maximum parsimony and Markov chain Monte Carlo Bayesian inference. MrBayes is [126] the most popular software for Bayesian inference. The Beast package can also be used for Bayesian inference [127]. Beast has a particular focus on calibrated phylogenies incorporating a time-scale [127], which was used to establish the phylogeny of four cone snail species based on the evolutionary history of gene duplications within a gene superfamily [19]. Maximum likelihood analyses of toxins were carried out using PhyML [128], PAUP* (Phylogenetic Analysis Using Parsimony) or MEGA (Molecular Evolutionary Genetics Analysis) [129]. MEGA was most often used for building maximum parsimony and neighbor joining trees in animal toxin studies. Substitution models were automatically selected using ModelGenerator [130], for protein and nucleic acid sequences, or ModelTest [131], for nucleic acid sequences only.

6.2. Evolutionary Analysis

Determining the positions under selective pressure could help in determining toxin functions. For example, positions under strong positive selections were shown to be important for selectivity and potency for some scorpion toxins [132]. CODEML of the PAML (Phylogenetic Analysis by Maximum Likelihood) package has often been employed to estimate several maximum-likelihood models that evaluate the ratio between non-synonymous to synonymous nucleotide substitution rates [133]. Another approach to compute this ratio is based on the Nei-Gojobori model [134], which is implemented in MEGA [129]. A ratio significantly greater than 1 indicates that the corresponding position is subject to positive selection (accelerated mutations, possibly for functional optimization); conversely, a ratio smaller than 1 indicates negative selection (slow mutation rates, possibly for maintaining function).

Fry and colleagues have employed Bayesian statistic methods to access better estimates of selective pressure on a range of toxin families [39,84,116] using BEB (Bayes Empirical Bayes) [135] and FUBAR (Fast, Unconstrained Bayesian AppRoximation) [136]. They also used an algorithm that works at the protein level called TreeSAAP [137] and a range of other algorithms to test episodic diversifying selection of toxins and taking into account the influence of recombination events [39,116].

7. Prediction of Toxin Structures and Activities

The functional characterization of toxins requires significant experimental effort, and consequently the majority of known toxins have not been experimentally tested for activity. A common assumption is that toxins belonging to the same gene superfamily and displaying the same cysteine framework will have similar targets. Whereas this approach often proves to be correct, some toxins sharing high sequence similarity have also shown different specificities. For example the spider toxins κ -theraphotoxin-Ps1b and β/ω -theraphotoxin-Tp2a share 77% sequence identity, but κ -theraphotoxin-Ps1b acts on voltage-gated potassium (K_v) and mechanosensitive channels [138,139], whereas β/ω -theraphotoxin-Tp2a is only active on voltage-gated sodium (Na_v) and Ca_v channels [140,141]. This section describes computational methods that could help assign or study toxin activity.

Toxins seem to adopt a limited number of structural scaffolds and it was initially hypothesized that the target selectivity of toxins could be predicted from statistical analysis of toxin sequences for each scaffold [22]. An algorithm based on neighbor-joining clustering with annotated scorpion toxins precursors was designed to predict ion channel targets, *i.e.*, specificity for Na_v , K_v , Ca_v or chloride channels [142], but this approach only resulted in 69% specificity. A more recent algorithm based on RBF network, which is a type of artificial neural network, achieved 89% accuracy at predicting the ion channel targeted by cone snail toxins [143]. This study used the frequency of di-peptide as features, and the most significant di-peptide sequences were selected based on the binomial distribution of their frequencies [144]. This approach offers hope that such bioinformatics approaches could accurately predict the type of targets of each animal toxin. Nevertheless, the prediction of toxin selectivity at the subtype level, which is the characteristic that generates most excitement about toxins, seems beyond reach using sequence information only. Indeed, a given toxin typically has a wide dynamic range of affinities, from picomolar to micromolar, on distinct molecular targets, tremendously complicating the prediction of toxin specificity through the sole use of amino acid sequence information. Predicting specificity for receptor subtypes requires detailed description of toxin molecular structures and interactions.

7.1. Molecular Modeling of Toxin Structures

Toxin structure-activity relationships studies aim to determine the structural features important for toxin target activity. Such studies proceed by comparing the 3D structures of wild-type and modified peptides to discover shared structural motifs that could be linked to function [115,122]. For example, cystine knot spider toxins that interact with the voltage sensors of ion channels display an important hydrophobic loop between their first two cystine residues [145] and this knowledge was used to suggest the function of newly discovered toxins [122]. Other structural features measured on structures are the accessible surface area of each residue and the electrostatic potential created by the toxins in a water environment, which could be computed by solving the Poisson–Boltzmann equation. The 3D structures of a large number of toxins have been solved by nuclear magnetic resonance (NMR), and these structures serve as templates to build homology models [39,116,122,146]. Most homology models were built using Modeller [147] or Swiss-Model (<http://swissmodel.expasy.org/>). In the absence of homologous toxins with experimentally determined structures, some studies attempted *ab initio* structural modeling. For example, the I-TASSER (Iterative Threading ASSEmby Refinement)

server [148] was used to model the scorpion toxin MeuTXK(beta)5-NHD(S) [115], and the PEP-FOLD server [149] was recently used to model the structure of cone snail toxin Vt3.1 [150]. I-TASSER provides a confidence score that can be used to assess if toxins at least adopt similar folds, and this score was used recently to suggest the fold adopted by three box jellyfish toxins [151]. In any modeling approach, it is important to remember that any given structural motif can be displayed by proteins having different activities; for example, the cystine knot scaffold is used by some ω -conotoxins to block calcium channels [152]; some κ -conotoxins to block potassium channels [153]; and some spider toxins to block sodium channels [154].

7.2. Molecular Modeling of Complexes with Molecular Targets

The molecular targets of toxins are typically very large transmembrane proteins, and there is currently only one crystallographic structure of a toxin in complex with its real molecular target: The tarantula psalmotoxin bound to a chicken acid-sensing ion channel [155]. Some structures of complexes between toxins and some structural surrogates have been determined experimentally [156]. Indeed, some cone snail and snake toxins have been co-crystallized with mollusk acetylcholine-binding proteins (AChBPs), which are structural homologs of the ligand-binding domain of nicotinic acetylcholine receptors (nAChRs) [157]. Molecular modeling is routinely used to gain insights on the molecular interactions at the origin of toxin affinity and specificity. The first step of this process is the structural modeling of the molecular targets, often performed by homology modeling based on some key experimental structures, including the crystal structure of rat Kv1.2 [158], the crystal structure of bacterial sodium channel NavAB [159], the electron microscopy structure of torpedo nAChR [160], the crystal structures of AChBPs [157] or the crystal structure of chicken ASIC1 channel [161]. Toxins often bind to a specific state of their molecular target, and the modeling of this state could be particularly difficult, especially if no relevant structural template is available, e.g., toxins binding to specific states of voltage-gated ion channels.

A determination of the binding mode of toxins binding to nAChR subtypes can confidently be modeled by homology on the basis of the conserved binding modes of homologs crystallized in complex with AChBP [162–167]. Some toxins are protease inhibitors and the modeling of their complexes, e.g., reference [168], can also be made confidently because of the large number of crystallographic structures of proteases. In the absence of structural homologs, molecular docking and different types of molecular dynamics (MD) simulations have been employed, as recently reviewed in detail [169]. The accuracy of such approaches greatly benefits from additional experimental data on the interaction, e.g., mutational data on both toxins and receptors, cross-linking experiments or interaction studies by NMR. Docking methods combine theoretical and heuristic techniques to produce a large number of docked “poses”, which are then clustered and ranked on the basis of a scoring function or using knowledge from experiments. The docking software Autodock [170], Gold (Statistical Innovations Inc., Belmont, MA, USA), Hex, ZDOCK [171] and HADDOCK [172] are among the docking algorithms the most frequently used for toxin binding studies. The ZDOCK scoring function allows some structural overlap at the interface, accounting for conformational variations at the interface. This method was successfully used to predict the interaction of mambalgin-2 toxin with ASIC1 channels, the fuzziness of the docking procedure allowing docking to the desensitized state even if the toxins is known to preferentially bind the closed state of the channel [173].

All-atom MD simulations are frequently employed to refine the conformation of models of molecular complex resulting from docking as well as to discover the binding modes of toxins. Because toxin/target systems typically have more than a hundred thousand atoms, the simulation times are limited to 10–100 ns, which is too short to observe spontaneous binding [169], except in some rare cases [174]. The use of biasing potentials (for example, harmonic restraints to maintain the toxin at proximity of the hypothesized binding site) allowed several studies to successfully propose some toxin binding modes [174,175]. An interesting and recent study used coarse-grained (CG) MD simulations instead of the standard all-atom MD to simulate the interactions between the spider toxin HaTx1 and a human Kv2.1 [176]. During a 3 μ s un-biased simulation, the HaTx1 first embedded in the phospholipid membrane bilayer and then interacted with the voltage-sensing domain of the channel [176]. CG representation substantially reduces the number of atoms in the simulated system, and simulations using CG force fields can consequently reach considerably longer time scales than all-atom MD simulations. Brownian dynamics is another type of simulation that has been used to predict the binding mode of several toxins without imposing any arbitrary restraints, and this technique was reviewed recently [169]. In this approach, the molecules are maintained partially rigid and water molecules are represented implicitly using the Generalized Born model, dramatically decreasing the number of atoms in the system [177]. Yu *et al.* [178] used Brownian dynamics to suggest that the six toxins AgTx2, ChTx, KTx, MgTx, NTx, and Pi2 have similar binding modes when bound to Kv1.3. The binding modes discovered by docking and MD could be experimentally validated by testing the activity of toxin variants, as suggested from the molecular models [163]. Another common approach to assess the validity of the resulting molecular model is to evaluate if it could be used to qualitatively explain mutation studies [164,167,174] or to reproduce absolute binding energies or relative mutational energies that have been measured experimentally [162,163,178].

7.3. Prediction of Toxin Binding Affinities and Specificities

As mentioned previously, the current major bottleneck in toxin research is the determination of the specificity of individual toxins, and several computational approaches that use extensive MD simulations to predict binding affinity could potentially help solve this problem. The most accurate method to quantitatively predict the binding affinity between peptides and proteins is the computation of potential of mean force (PMF) along unbinding paths sampled using umbrella sampling MD simulations [179]. Two recent reviews present, in detail, the use of this technique for studying the binding of animal toxins to ion channels [169,180]. The umbrella sampling procedure consists of carrying out a series of MD simulations starting at different coordinates along the unbinding pathway. The average energy as a function of the reaction coordinate, *i.e.*, the PMF, is then computed for each simulation, and the PMFs are finally combined into one using the weighted histogram analysis method (WHAM) [181]. In a recent study using this technique, the predicted dissociation constants of the toxins Csx4 and Cn2 from a Nav were 20 nM and 70 nM, which are comparable to the experimental dissociation constants of 1 nM and 40 nM, respectively [182]. In general, the dissociation constants predicted using the PMF/umbrella sampling method are within one or two orders of magnitude of experimental values [169]. An alternative technique used to build the PMF consists of using the Jarzynski equation to analyze a simulation in which unbinding pathways are

rapidly sampled using a pulling potential [183]. When applied to a system comprising a toxin and a potassium channel, the method was found to be less sensitive and required more simulations to converge than the weighted histogram method [184]. Another study that focused on a complex between conotoxin ImI and a nAChR used a slower pulling speed and obtained at reasonable computational cost a one fold difference between the predicted (-70 kcal/mol) and experimental (-9 kcal/mol) binding free energies [164].

The PMF-based methods are computationally intensive, and other less accurate methods need to be employed to rapidly screen large libraries of toxins. One popular approximation requires sampling the conformational ensembles in the unbound and bound states only, and the binding energy is approximated by the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) function [185]. This method was employed to study a system comprising conotoxin ImI and nAChR $\alpha 7$, and the predicted mutational energies of 16 ImI variants correlated with experimental values with a R^2 correlation coefficient of 0.74, which is considered as accurate [162]. MM/PBSA was also successfully employed in several other *in silico* mutational studies of toxin/target systems [163,186]. For example, it was recently used to determine the binding sites of the analgesic conotoxin Vc1.1 on the $\alpha 9\alpha 10$ nAChR subtype, quantitatively predicting the impact of mutations that were later used to validate the binding mode experimentally [163]. The MM/PBSA method has a low computational cost but can potentially produce large errors depending of the simulation protocol [187]. We anticipate that this type of approach could be used to predict the subtype specificity of animal toxins, even if the accurate predictions of binding affinities are beyond its scope.

8. Future Perspective: Integrating the Biology of Venoms and Prediction of Toxin Activity

Bioinformatics is important at every step of venomics-based discovery. Databases orient venomics research by providing a concise overview of current knowledge on toxins, highlighting the venomous species or the type of toxins that have higher potential to become drugs, and suggesting what venomics fields are still terra incognita. In recent years, specific approaches have been invented to mine combined data from transcriptomics and proteomics, dramatically accelerating the pace of toxin discovery. These new approaches trade accuracy for speed, and we expect future improvements of the quality of experimental data as well as of the reliability assessment of results produced by venomics specific software. Our knowledge of venoms has rapidly expanded, and our estimate of the number of toxins has increased even more rapidly. We indeed now realize that venom biology is important and complex, as exemplified by some observations of extreme intra-specimen venom variability. We now also comprehend the complexity of venom chemistry, especially since it is now clear that venoms contain many more peptide toxins than there are genes encoding toxins in the genome. Currently, the description of these types of diversities is not well managed by venomics databases and bioinformatics tools. The aim of venomics efforts is to rapidly discover peptides with pharmaceutical activity, but paradoxically the venomics “pipeline” does not incorporate any new experimental approach to rapidly assess peptide functions. Computational approaches to predict the function of toxins already exist but applying them to the study of all toxins is not practical. Indeed, the most accurate computational methods to predict toxin affinities require a computation time higher

than measuring activity by electrophysiology. Venomics is in great need of new molecular modeling and bioinformatics tools able to suggest the molecular target specificity of toxins.

Acknowledgments

This project was supported by an Australian Research Council grant (DP150103990) and by an Early Career Research (ECR) grant from the University of Queensland. D.J.C. is a National Health and Medical Research Council Professorial Fellow (APP1026501).

Author Contributions

Q.K. and D.J.C. wrote the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Georgieva, D.; Arni, R.K.; Betzel, C. Proteome analysis of snake venom toxins: Pharmacological insights. *Expert Rev. Proteomics* **2008**, *5*, 787–797.
2. Escoubas, P.; Quinton, L.; Nicholson, G.M. Venomics: Unravelling the complexity of animal venoms with mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 279–295.
3. Davis, J.; Jones, A.; Lewis, R.J. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* **2009**, *30*, 1222–1227.
4. Duda, T.F., Jr.; Kohn, A.J.; Matheny, A.M. Cryptic species differentiated in *Conus ebraeus*, a widespread tropical marine gastropod. *Biol. Bull.* **2009**, *217*, 292–305.
5. Puillandre, N.; Bouchet, P.; Duda, T.F.; Kaufenstein, S.; Kohn, A.J.; Olivera, B.M.; Watkins, M.; Meyer, C. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol. Phylogenet. Evol.* **2014**, *78*, 290–303.
6. World Spider Catalog. Available online: <http://www.wsc.nmbe.ch/> (accessed on 5 June 2015).
7. Vidal, N.; Delmas, A.-S.; David, P.; Cruaud, C.; Couloux, A.; Hedges, S.B. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *C. R. Biol.* **2007**, *330*, 182–187.
8. Uetz, P. The Reptile Database. Available online: <http://www.reptile-database.org> (accessed on 5 June 2015).
9. Cao, Z.; Yu, Y.; Wu, Y.; Hao, P.; Di, Z.; He, Y.; Chen, Z.; Yang, W.; Shen, Z.; He, X.; *et al.* The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat. Commun.* **2013**, *4*, 2602.
10. Calvete, J.J. Venomics, what else? *Toxicon* **2012**, *60*, 427–433.
11. Miljanich, G.P. Ziconotide: Neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.* **2004**, *11*, 3029–3040.
12. Beraud, E.; Chandy, K.G. Therapeutic potential of peptide toxins that target ion channels. *Inflamm. Allergy Drug Targets* **2011**, *10*, 322–342.

13. Lewis, R.J.; Dutertre, S.; Vetter, I.; Christie, M.J. Conus venom peptide pharmacology. *Pharmacol. Rev.* **2012**, *64*, 259–298.
14. Mebs, D. Toxicity in animals. Trends in evolution? *Toxicon* **2001**, *39*, 87–96.
15. Kaas, Q.; Westermann, J.-C.; Craik, D.J. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon* **2010**, *55*, 1491–1509.
16. 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.
17. Buczek, O.; Bulaj, G.; Olivera, B.M. Conotoxins and the posttranslational modification of secreted gene products. *Cell. Mol. Life Sci.* **2005**, *62*, 3067–3079.
18. Akondi, K.B.; Muttenthaler, M.; Dutertre, S.; Kaas, Q.; Craik, D.J.; Lewis, R.J.; Alewood, P.F. Discovery, synthesis, and structure-activity relationships of conotoxins. *Chem. Rev.* **2014**, *114*, 5815–5847.
19. 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.
20. Prashanth, J.R.; Lewis, R.J.; Dutertre, S. Towards an integrated venomomics approach for accelerated conopeptide discovery. *Toxicon* **2012**, *60*, 470–477.
21. Juárez, P.; Sanz, L.; Calvete, J.J. Snake venomomics: Characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* **2004**, *4*, 327–338.
22. Tan, P.T.J.; Khan, A.M.; Brusica, V. Bioinformatics for venom and toxin sciences. *Brief. Bioinform.* **2003**, *4*, 53–62.
23. Benson, D.A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2014**, *42*, D32–D37.
24. Jungo, F.; Estreicher, A.; Bairoch, A.; Bougueleret, L.; Xenarios, I. Animal toxins: How is complexity represented in databases? *Toxins* **2010**, *2*, 262–282.
25. 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.* **2012**, *40*, D325–D330.
26. Herzig, V.; Wood, D.L.A.; Newell, F.; Chaumeil, P.-A.; Kaas, Q.; Binford, G.J.; Nicholson, G.M.; Gorse, D.; King, G.F. ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures. *Nucleic Acids Res.* **2011**, *39*, D653–D657.
27. Roly, Z.Y.; Hakim, M.A.; Zahan, A.S.; Hossain, M.M.; Reza, M.A. ISOB: A Database of Indigenous Snake Species of Bangladesh with respective known venom composition. *Bioinformation* **2015**, *11*, 107–114.
28. The UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* **2014**, *42*, D191–D198.
29. Gutmanas, A.; Alhroub, Y.; Battle, G.M.; Berrisford, J.M.; Bochet, E.; Conroy, M.J.; Dana, J.M.; Fernandez Montecelo, M.A.; van Ginkel, G.; Gore, S.P.; *et al.* PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* **2014**, *42*, D285–D291.
30. King, G.F.; Gentz, M.C.; Escoubas, P.; Nicholson, G.M. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* **2008**, *52*, 264–276.
31. Kaas, Q.; Westermann, J.-C.; Halai, R.; Wang, C.K.L.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinform. Oxf. Engl.* **2008**, *24*, 445–446.

32. Wood, D.L.A.; Miljenović, T.; Cai, S.; Raven, R.J.; Kaas, Q.; Escoubas, P.; Herzig, V.; Wilson, D.; King, G.F. ArachnoServer: A database of protein toxins from spiders. *BMC Genomics* **2009**, *10*, 375.
33. Jungo, F.; Bairoch, A. Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase. *Toxicon* **2005**, *45*, 293–301.
34. He, Q.; Han, W.; He, Q.; Huo, L.; Zhang, J.; Lin, Y.; Chen, P.; Liang, S. ATDB 2.0: A database integrated toxin-ion channel interaction data. *Toxicon* **2010**, *56*, 644–647.
35. Pawson, A.J.; Sharman, J.L.; Benson, H.E.; Faccenda, E.; Alexander, S.P.H.; Buneman, O.P.; Davenport, A.P.; McGrath, J.C.; Peters, J.A.; Southan, C.; *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY: An expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.* **2014**, *42*, D1098–D1106.
36. Markley, J.L.; Ulrich, E.L.; Berman, H.M.; Henrick, K.; Nakamura, H.; Akutsu, H. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): New policies affecting biomolecular NMR depositions. *J. Biomol. NMR* **2008**, *40*, 153–155.
37. Rawlings, N.D.; Waller, M.; Barrett, A.J.; Bateman, A. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014**, *42*, D503–D509.
38. Tan, P.T.J.; Veeramani, A.; Srinivasan, K.N.; Ranganathan, S.; Brusica, V. SCORPION2: A database for structure-function analysis of scorpion toxins. *Toxicon* **2006**, *47*, 356–363.
39. Sunagar, K.; Undheim, E.A.B.; Chan, A.H.C.; Koludarov, I.; Muñoz-Gómez, S.A.; Antunes, A.; Fry, B.G. Evolution stings: The origin and diversification of scorpion toxin peptide scaffolds. *Toxins* **2013**, *5*, 2456–2487.
40. Duda, T.F.; Chang, D.; Lewis, B.D.; Lee, T. Geographic variation in venom allelic composition and diets of the widespread predatory marine gastropod *Conus ebraeus*. *PLoS ONE* **2009**, doi:10.1371/journal.pone.0006245.
41. Castoe, T.A.; de Koning, A.P.J.; Hall, K.T.; Card, D.C.; Schield, D.R.; Fujita, M.K.; Ruggiero, R.P.; Degner, J.F.; Daza, J.M.; Gu, W.; *et al.* The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20645–20650.
42. Vonk, F.J.; Casewell, N.R.; Henkel, C.V.; Heimberg, A.M.; Jansen, H.J.; McCleary, R.J.R.; Kerkkamp, H.M.E.; Vos, R.A.; Guerreiro, I.; Calvete, J.J.; *et al.* The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20651–20656.
43. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics* **2011**, *12*, 60.
44. 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.
45. Sanggaard, K.W.; Bechsgaard, J.S.; Fang, X.; Duan, J.; Dyrland, T.F.; Gupta, V.; Jiang, X.; Cheng, L.; Fan, D.; Feng, Y.; *et al.* Spider genomes provide insight into composition and evolution of venom and silk. *Nat. Commun.* **2014**, *5*, 3765.
46. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **2006**, *443*, 931–949.

47. Werren, J.H.; Richards, S.; Desjardins, C.A.; Niehuis, O.; Gadau, J.; Colbourne, J.K.; Nasonia Genome Working Group; Werren, J.H.; Richards, S.; Desjardins, C.A.; *et al.* Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **2010**, *327*, 343–348.
48. Nagaraj, S.H.; Gasser, R.B.; Ranganathan, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief. Bioinform.* **2007**, *8*, 6–21.
49. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology information. *Nucleic Acids Res.* **2014**, *42*, D7–D17.
50. Jin, A.; 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. Proteomics* **2013**, *12*, 3824–3833.
51. Gilles, A.; Megléc, E.; Pech, N.; Ferreira, S.; Malausa, T.; Martin, J.-F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **2011**, *12*, 245.
52. Lavergne, V.; Dutertre, S.; Jin, A.; 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 Genomics* **2013**, *14*, 708.
53. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999**, *9*, 868–877.
54. Gordon, D.; Green, P. Consed: A graphical editor for next-generation sequencing. *Bioinformatics* **2013**, *29*, 2936–2937.
55. Swindell, S.R.; Plasterer, T.N. SEQMAN. Contig assembly. *Methods Mol. Biol.* **1997**, *70*, 75–89.
56. Chevreux, B.; Pfisterer, T.; Drescher, B.; Driesel, A.J.; Müller, W.E.G.; Wetter, T.; Suhai, S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **2004**, *14*, 1147–1159.
57. Schulz, M.H.; Zerbino, D.R.; Vingron, M.; Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **2012**, *28*, 1086–1092.
58. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652.
59. Gorson, J.; Ramrattan, G.; Verdes, A.; Wright, M.E.; Kantor, Y.; Srinivasan, R.; Musunuri, R.; Packer, D.; Albano, G.; Qiu, W.G.; *et al.* Molecular diversity and gene evolution of the venom arsenal of terebridae predatory marine snails. *Genome Biol. Evol.* **2015**, doi:10.1093/gbe/evv104.
60. Dutertre, S.; Jin, A.; Kaas, Q.; Jones, A.; Alewood, P.; Lewis, R. Deep venomomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* **2013**, *12*, 312–329.
61. Archer, J.; Whiteley, G.; Casewell, N.R.; Harrison, R.A.; Wagstaff, S.C. VTBuilder: A tool for the assembly of multi isoform transcriptomes. *BMC Bioinform.* **2014**, *15*, 389.
62. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676.
63. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240.

64. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185.
65. Yamada, T.; Letunic, I.; Okuda, S.; Kanehisa, M.; Bork, P. iPath2.0: Interactive pathway explorer. *Nucleic Acids Res.* **2011**, *39*, W412–W415.
66. 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 Genomics* **2012**, *13*, 284.
67. 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**, doi:10.1371/journal.pone.0087648.
68. Schwartz, E.F.; Diego-Garcia, E.; de la Vega, R.C.R.; Possani, L.D. Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC Genomics* **2007**, *8*, 119.
69. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
70. 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.
71. 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. Biophys. Acta* **2013**, *1834*, 717–724.
72. Petrel, C.; Hocking, H.G.; Reynaud, M.; Upert, G.; Favreau, P.; Biass, D.; Paolini-Bertrand, M.; Peigneur, S.; Tytgat, J.; Gilles, N.; *et al.* Identification, structural and pharmacological characterization of τ -CnVA, a conopeptide that selectively interacts with somatostatin sst3 receptor. *Biochem. Pharmacol.* **2013**, *85*, 1663–1671.
73. Eddy, S.R. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput. Biol.* **2008**, *4*, e1000069.
74. Kozlov, S.; Malyavka, A.; McCutchen, B.; Lu, A.; Schepers, E.; Herrmann, R.; Grishin, E. A novel strategy for the identification of toxinlike structures in spider venom. *Proteins* **2005**, *59*, 131–140.
75. Gracy, J.; Le-Nguyen, D.; Gelly, J.-C.; Kaas, Q.; Heitz, A.; Chiche, L. KNOTTIN: The knottin or inhibitor cystine knot scaffold in 2007. *Nucleic Acids Res.* **2008**, *36*, D314–D319.
76. Torres, A.F.C.; Huang, C.; Chong, C.-M.; Leung, S.W.; Prieto-da-Silva, Á.R.B.; Havt, A.; Quinet, Y.P.; Martins, A.M.C.; Lee, S.M.Y.; Rádis-Baptista, G. Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadriceps*: Insights into the polypeptide toxin arsenal of hymenopterans. *PLoS ONE* **2014**, doi:10.1371/journal.pone.0087556.
77. Naamati, G.; Askenazi, M.; Linal, M. ClanTox: A classifier of short animal toxins. *Nucleic Acids Res.* **2009**, *37*, W363–W368.
78. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
79. Tirosh, Y.; Ofer, D.; Eliyahu, T.; Linal, M. Short toxin-like proteins attack the defense line of innate immunity. *Toxins* **2013**, *5*, 1314–1331.

80. Lyukmanova, E.N.; Shulepko, M.A.; Buldakova, S.L.; Kasheverov, I.E.; Shenkarev, Z.O.; Reshetnikov, R.V.; Filkin, S.Y.; Kudryavtsev, D.S.; *et al.* Water-soluble LYNX1 residues important for interaction with muscle-type and/or neuronal nicotinic receptors. *J. Biol. Chem.* **2013**, *288*, 15888–15899.
81. Shulepko, M.A.; Lyukmanova, E.N.; Paramonov, A.S.; Lobas, A.A.; Shenkarev, Z.O.; Kasheverov, I.E.; Dolgikh, D.A.; Tsetlin, V.I.; Arseniev, A.S.; Kirpichnikov, M.P. Human neuromodulator SLURP-1: Bacterial expression, binding to muscle-type nicotinic acetylcholine receptor, secondary structure, and conformational heterogeneity in solution. *Biochemistry* **2013**, *78*, 204–211.
82. Starcevic, A.; Moura-da-Silva, A.M.; Cullum, J.; Hranueli, D.; Long, P.F. Combinations of long peptide sequence blocks can be used to describe toxin diversification in venomous animals. *Toxicon* **2015**, *95*, 84–92.
83. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208.
84. 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.
85. Inceoglu, B.; Lango, J.; Jing, J.; Chen, L.; Doymaz, F.; Pessah, I.N.; Hammock, B.D. One scorpion, two venoms: Prevenom of *Parabuthus transvaalicus* acts as an alternative type of venom with distinct mechanism of action. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 922–927.
86. Franco, A.; Pisarewicz, K.; Moller, C.; Mora, D.; Fields, G.B.; Mari, F. Hyperhydroxylation: A new strategy for neuronal targeting by venomous marine molluscs. *Prog. Mol. Subcell. Biol.* **2006**, *43*, 83–103.
87. Tawfik, D.S. Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* **2010**, *6*, 692–696.
88. Gordon, A.J.E.; Satory, D.; Halliday, J.A.; Herman, C. Lost in transcription: Transient errors in information transfer. *Curr. Opin. Microbiol.* **2015**, *24*, 80–87.
89. Gout, J.-F.; Thomas, W.K.; Smith, Z.; Okamoto, K.; Lynch, M. Large-scale detection of *in vivo* transcription errors. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 18584–18589.
90. Fox, J.W.; Serrano, S.M.T. Exploring snake venom proteomes: Multifaceted analyses for complex toxin mixtures. *Proteomics* **2008**, *8*, 909–920.
91. De Queiroz, M.R.; Mamede, C.C.N.; de Moraes, N.C.; Fonseca, K.C.; de Sousa, B.B.; Migliorini, T.M.; Pereira, D.F.; Stanziola, L.; Calderon, L.A.; Simões-Silva, R.; *et al.* Purification and Characterization of BmooAi: A New Toxin from *Bothrops moojeni* Snake Venom That Inhibits Platelet Aggregation. *BioMed Res. Int.* **2014**, *2014*, doi:10.1155/2014/920942.
92. Hardy, M.C.; Daly, N.L.; Mobli, M.; Morales, R.A.V.; King, G.F. Isolation of an orally active insecticidal toxin from the venom of an Australian tarantula. *PLoS ONE* **2013**, doi:10.1371/journal.pone.0073136.
93. Calvete, J.J.; Ghezellou, P.; Paiva, O.; Matainaho, T.; Ghassempour, A.; Goudarzi, H.; Kraus, F.; Sanz, L.; Williams, D.J. Snake venomomics of two poorly known Hydrophiinae: Comparative proteomics of the venoms of terrestrial *Toxicocalamus longissimus* and marine *Hydrophis cyanocinctus*. *J. Proteomics* **2012**, *75*, 4091–4101.

94. Bhatia, S.; Kil, Y.J.; Ueberheide, B.; Chait, B.T.; Tayo, L.; Cruz, L.; Lu, B.; Yates, J.R., 3rd; Bern, M. Constrained *de novo* sequencing of conotoxins. *J. Proteome Res.* **2012**, *11*, 4191–4200.
95. Escoubas, P.; Sollod, B.; King, G.F. Venom landscapes: Mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* **2006**, *47*, 650–663.
96. Craig, R.; Beavis, R.C. TANDEM: Matching proteins with tandem mass spectra. *Bioinform. Oxf. Engl.* **2004**, *20*, 1466–1467.
97. Hoopmann, M.R.; Moritz, R.L. Current algorithmic solutions for peptide-based proteomics data generation and identification. *Curr. Opin. Biotechnol.* **2013**, *24*, 31–38.
98. Bunkenborg, J.; Matthiesen, R. Interpretation of tandem mass spectra of posttranslationally modified peptides. *Methods Mol. Biol.* **2013**, *1007*, 139–171.
99. Skjaerbaek, N.; Nielsen, K.J.; Lewis, R.J.; Alewood, P.; Craik, D.J. Determination of the solution structures of conantokin-G and conantokin-T by CD and NMR spectroscopy. *J. Biol. Chem.* **1997**, *272*, 2291–2299.
100. Dutton, J.L.; Bansal, P.S.; Hogg, R.C.; Adams, D.J.; Alewood, P.F.; Craik, D.J. A new level of conotoxin diversity, a non-native disulfide bond connectivity in α -conotoxin Au1B reduces structural definition but increases biological activity. *J. Biol. Chem.* **2002**, *277*, 48849–48857.
101. Wilkins, M.R.; Gasteiger, E.; Gooley, A.A.; Herbert, B.R.; Molloy, M.P.; Binz, P.A.; Ou, K.; Sanchez, J.C.; Bairoch, A.; Williams, K.L.; *et al.* High-throughput mass spectrometric discovery of protein post-translational modifications. *J. Mol. Biol.* **1999**, *289*, 645–657.
102. Pang, C.N.I.; Gasteiger, E.; Wilkins, M.R. Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. *BMC Genomics* **2010**, *11*, 92.
103. Calvete, J.J.; Marcinkiewicz, C.; Sanz, L. Snake venomomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. *J. Proteome Res.* **2007**, *6*, 326–336.
104. Durban, J.; Pérez, A.; Sanz, L.; Gómez, A.; Bonilla, F.; Rodríguez, S.; Chacón, D.; Sasa, M.; Angulo, Y.; Gutiérrez, J.M.; *et al.* Integrated “omics” profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. *BMC Genomics* **2013**, *14*, 234.
105. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Petersen* **2011**, *8*, 785–786.
106. Kozlov, S.A.; Grishin, E.V. The universal algorithm of maturation for secretory and excretory protein precursors. *Toxicon* **2007**, *49*, 721–726.
107. Luo, S.; Zhangsun, D.; Wu, Y.; Zhu, X.; Xie, L.; Hu, Y.; Zhang, J.; Zhao, X. Identification and molecular diversity of T-superfamily conotoxins from *Conus lividus* and *Conus litteratus*. *Chem. Biol. Drug Des.* **2006**, *68*, 97–106.
108. Duckert, P.; Brunak, S.; Blom, N. Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.* **2004**, *17*, 107–112.
109. Wong, E.S.W.; Hardy, M.C.; Wood, D.; Bailey, T.; King, G.F. SVM-based prediction of propeptide cleavage sites in spider toxins identifies toxin innovation in an Australian tarantula. *PLoS ONE* **2013**, doi:10.1371/journal.pone.0066279.

110. Milne, T.J.; Abbenante, G.; Tyndall, J.D.A.; Halliday, J.; Lewis, R.J. Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J. Biol. Chem.* **2003**, *278*, 31105–31110.
111. 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.
112. 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.
113. 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.
114. 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.
115. Zhu, S.; Peigneur, S.; Gao, B.; Luo, L.; Jin, D.; Zhao, Y.; Tytgat, J. Molecular diversity and functional evolution of scorpion potassium channel toxins. *Mol. Cell. Proteomics* **2011**, *10*, doi:10.1074/mcp.M110.002832.
116. Sunagar, K.; Jackson, T.N.W.; Undheim, E.A.B.; Ali, S.A.; Antunes, A.; Fry, B.G. Three-fingered RAVERS: Rapid Accumulation of Variations in Exposed Residues of snake venom toxins. *Toxins* **2013**, *5*, 2172–2208.
117. Fry, B.G.; Undheim, E.A.B.; Ali, S.A.; Jackson, T.N.W.; Debono, J.; Scheib, H.; Ruder, T.; Morgenstern, D.; Cadwallader, L.; Whitehead, D.; *et al.* Squeezers and leaf-cutters: Differential diversification and degeneration of the venom system in toxicoferan reptiles. *Mol. Cell. Proteomics* **2013**, *12*, 1881–1899.
118. Olivera, B. *Conus venom* peptides: Reflections from the biology of clades and species. *Annu. Rev. Ecol. Syst.* **2002**, *33*, 25–47.
119. 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. Phylogenet. Evol.* **2010**, *56*, 1–12.
120. Fry, B.G.; Wüster, W.; Kini, R.M.; Brusich, V.; Khan, A.; Venkataraman, D.; Rooney, A.P. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* **2003**, *57*, 110–129.
121. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
122. Chen, J.; Deng, M.; He, Q.; Meng, E.; Jiang, L.; Liao, Z.; Rong, M.; Liang, S. Molecular diversity and evolution of cystine knot toxins of the tarantula *Chilobrachys jingzhao*. *Cell. Mol. Life Sci.* **2008**, *65*, 2431–2444.
123. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113.
124. Puillandre, N.; Koua, D.; Favreau, P.; Olivera, B.M.; Stöcklin, R. Molecular phylogeny, classification and evolution of conopeptides. *J. Mol. Evol.* **2012**, *74*, 297–309.

125. Duda, T.F., Jr.; Remigio, E.A. Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. *Mol. Ecol.* **2008**, *17*, 3018–3032.
126. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542.
127. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2014**, doi:10.1371/journal.pcbi.1003537.
128. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321.
129. Tamura, K.; Stecher, G.; Peterson, D.; Filipowski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729.
130. Keane, T.M.; Creevey, C.J.; Pentony, M.M.; Naughton, T.J.; McInerney, J.O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **2006**, *6*, 29.
131. Posada, D.; Crandall, K.A. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **1998**, *14*, 817–818.
132. Weinberger, H.; Moran, Y.; Gordon, D.; Turkov, M.; Kahn, R.; Gurevitz, M. Positions under positive selection—Key for selectivity and potency of scorpion α -toxins. *Mol. Biol. Evol.* **2010**, *27*, 1025–1034.
133. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591.
134. Zhang, J.; Rosenberg, H.F.; Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3708–3713.
135. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118.
136. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Pond, S.L.K.; Scheffler, K. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205.
137. Woolley, S.; Johnson, J.; Smith, M.J.; Crandall, K.A.; McClellan, D.A. TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinform. Oxf. Engl.* **2003**, *19*, 671–672.
138. Diochot, S.; Drici, M.D.; Moinier, D.; Fink, M.; Lazdunski, M. Effects of phrixotoxins on the Kv4 family of potassium channels and implications for the role of Ito1 in cardiac electrogenesis. *Br. J. Pharmacol.* **1999**, *126*, 251–263.
139. Oswald, R.E.; Suchyna, T.M.; McFeeters, R.; Gottlieb, P.; Sachs, F. Solution structure of peptide toxins that block mechanosensitive ion channels. *J. Biol. Chem.* **2002**, *277*, 34443–34450.
140. Edgerton, G.B.; Blumenthal, K.M.; Hanck, D.A. Inhibition of the activation pathway of the T-type calcium channel Ca(V)3.1 by ProTxII. *Toxicon* **2010**, *56*, 624–636.
141. Middleton, R.E.; Warren, V.A.; Kraus, R.L.; Hwang, J.C.; Liu, C.J.; Dai, G.; Brochu, R.M.; Kohler, M.G.; Gao, Y.-D.; Garsky, V.M.; *et al.* Two tarantula peptides inhibit activation of multiple sodium channels. *Biochemistry* **2002**, *41*, 14734–14747.

142. Tan, P.T.J.; Srinivasan, K.N.; Seah, S.H.; Koh, J.L.Y.; Tan, T.W.; Ranganathan, S.; Brusica, V. Accurate prediction of scorpion toxin functional properties from primary structures. *J. Mol. Graph. Model.* **2005**, *24*, 17–24.
143. 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. In Vitro* **2013**, *27*, 852–856.
144. Feng, Y.; Luo, L. Use of tetrapeptide signals for protein secondary-structure prediction. *Amino Acids* **2008**, *35*, 607–614.
145. Lee, S.-Y.; MacKinnon, R. A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature* **2004**, *430*, 232–235.
146. Almeida, D.D.; Torres, T.M.; Barbosa, E.G.; Lima, J.P.M.S.; de Freitas Fernandes-Pedrosa, M. Molecular approaches for structural characterization of a new potassium channel blocker from *Tityus stigmurus* venom: cDNA cloning, homology modeling, dynamic simulations and docking. *Biochem. Biophys. Res. Commun.* **2013**, *430*, 113–118.
147. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
148. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738.
149. Maupetit, J.; Derreumaux, P.; Tuffery, P. PEP-FOLD: An online resource for *de novo* peptide structure prediction. *Nucleic Acids Res.* **2009**, *37*, W498–W503.
150. Li, M.; Chang, S.; Yang, L.; Shi, J.; McFarland, K.; Yang, X.; Moller, A.; Wang, C.; Zou, X.; Chi, C.; *et al.* Conopeptide Vt3.1 preferentially inhibits BK potassium channels containing $\beta 4$ subunits via electrostatic interactions. *J. Biol. Chem.* **2014**, *289*, 4735–4742.
151. Brinkman, D.L.; Konstantakopoulos, N.; McInerney, B.V.; Mulvenna, J.; Seymour, J.E.; Isbister, G.K.; Hodgson, W.C. *Chironex fleckeri* (box jellyfish) venom proteins: Expansion of a cnidarian toxin family that elicits variable cytolytic and cardiovascular effects. *J. Biol. Chem.* **2014**, *289*, 4798–4812.
152. Nielsen, K.J.; Thomas, L.; Lewis, R.J.; Alewood, P.F.; Craik, D.J. A consensus structure for omega-conotoxins with different selectivities for voltage-sensitive calcium channel subtypes: Comparison of MVIIA, SVIB and SNX-202. *J. Mol. Biol.* **1996**, *263*, 297–310.
153. Scanlon, M.J.; Naranjo, D.; Thomas, L.; Alewood, P.F.; Lewis, R.J.; Craik, D.J. Solution structure and proposed binding mechanism of a novel potassium channel toxin κ -conotoxin PVIIA. *Structure* **1997**, *5*, 1585–1597.
154. Fletcher, J.I.; Chapman, B.E.; Mackay, J.P.; Howden, M.E.; King, G.F. The structure of versutoxin (δ -atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel. *Structure* **1997**, *5*, 1525–1535.
155. Bacongus, I.; Gouaux, E. Structural plasticity and dynamic selectivity of acid-sensing ion channel-spider toxin complexes. *Nature* **2012**, *489*, 400–405.
156. Dutertre, S.; Ulens, C.; Büttner, R.; Fish, A.; van Elk, R.; Kendel, Y.; Hopping, G.; Alewood, P.F.; Schroeder, C.; Nicke, A.; *et al.* AChBP-targeted α -conotoxin correlates distinct binding orientations with nAChR subtype selectivity. *EMBO J.* **2007**, *26*, 3858–3867.

157. Tsetlin, V.; Utkin, Y.; Kasheverov, I. Polypeptide and peptide toxins, magnifying lenses for binding sites in nicotinic acetylcholine receptors. *Biochem. Pharmacol.* **2009**, *78*, 720–731.
158. Long, S.B.; Campbell, E.B.; MacKinnon, R. Voltage sensor of Kv1.2: Structural basis of electromechanical coupling. *Science* **2005**, *309*, 903–908.
159. Payandeh, J.; Scheuer, T.; Zheng, N.; Catterall, W.A. The crystal structure of a voltage-gated sodium channel. *Nature* **2011**, *475*, 353–358.
160. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. *J. Mol. Biol.* **2005**, *346*, 967–989.
161. Jasti, J.; Furukawa, H.; Gonzales, E.B.; Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* **2007**, *449*, 316–323.
162. Yu, R.; Craik, D.J.; Kaas, Q. Blockade of neuronal $\alpha 7$ -nAChR by α -conotoxin ImI explained by computational scanning and energy calculations. *PLoS Comput. Biol.* **2011**, doi:10.1371/journal.pcbi.1002011.
163. Yu, R.; Kompella, S.N.; Adams, D.J.; Craik, D.J.; Kaas, Q. Determination of the α -conotoxin Vc1.1 binding site on the $\alpha 9\alpha 10$ nicotinic acetylcholine receptor. *J. Med. Chem.* **2013**, *56*, 3557–3567.
164. Yu, R.; Kaas, Q.; Craik, D.J. Delineation of the unbinding pathway of α -conotoxin ImI from the $\alpha 7$ nicotinic acetylcholine receptor. *J. Phys. Chem. B* **2012**, *116*, 6097–6105.
165. Luo, S.; Zhangsun, D.; Schroeder, C.I.; Zhu, X.; Hu, Y.; Wu, Y.; Weltzin, M.M.; Eberhard, S.; Kaas, Q.; Craik, D.J.; *et al.* A novel $\alpha 4/7$ -conotoxin LvIA from *Conus lividus* that selectively blocks $\alpha 3\beta 2$ vs. $\alpha 6/\alpha 3\beta 2\beta 3$ nicotinic acetylcholine receptors. *FASEB J.* **2014**, *28*, 1842–1853.
166. Grishin, A.A.; Cuny, H.; Hung, A.; Clark, R.J.; Brust, A.; Akondi, K.; Alewood, P.F.; Craik, D.J.; Adams, D.J. Identifying key amino acid residues that affect α -conotoxin AuIB inhibition of $\alpha 3\beta 4$ nicotinic acetylcholine receptors. *J. Biol. Chem.* **2013**, *288*, 34428–34442.
167. Pucci, L.; Grazioso, G.; Dallanoce, C.; Rizzi, L.; de Micheli, C.; Clementi, F.; Bertrand, S.; Bertrand, D.; Longhi, R.; de Amici, M.; *et al.* Engineering of α -conotoxin MII-derived peptides with increased selectivity for native $\alpha 6\beta 2^*$ nicotinic acetylcholine receptors. *FASEB J.* **2011**, *25*, 3775–3789.
168. Liu, H.; Chen, J.; Wang, X.; Yan, S.; Xu, Y.; San, M.; Tang, W.; Yang, F.; Cao, Z.; Li, W.; *et al.* Functional characterization of a new non-Kunitz serine protease inhibitor from the scorpion *Lychas mucronatus*. *Int. J. Biol. Macromol.* **2015**, *72C*, 158–162.
169. Gordon, D.; Chen, R.; Chung, S.-H. Computational methods of studying the binding of toxins from venomous animals to biological ion channels: Theory and applications. *Physiol. Rev.* **2013**, *93*, 767–802.
170. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
171. Pierce, B.G.; Hourai, Y.; Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* **2011**, doi:10.1371/journal.pone.0024657.
172. Kastritis, P.L.; Rodrigues, J.P.G.L.M.; Bonvin, A.M.J.J. HADDOCK(2P2I): A biophysical model for predicting the binding affinity of protein-protein interaction inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 826–836.

173. Salinas, M.; Besson, T.; Delettre, Q.; Diochot, S.; Boulakirba, S.; Douguet, D.; Lingueglia, E. Binding site and inhibitory mechanism of the mambalgin-2 pain-relieving peptide on acid-sensing ion channel 1a. *J. Biol. Chem.* **2014**, *289*, 13363–13373.
174. Chen, R.; Chung, S.-H. Binding modes and functional surface of anti-mammalian scorpion α -toxins to sodium channels. *Biochemistry* **2012**, *51*, 7775–7782.
175. Chen, R.; Chung, S.-H. Structural basis of the selective block of Kv1.2 by maurotoxin from computer simulations. *PLoS ONE* **2012**, doi:10.1371/journal.pone.0047253.
176. Wee, C.L.; Gavaghan, D.; Sansom, M.S.P. Interactions between a voltage sensor and a toxin via multiscale simulations. *Biophys. J.* **2010**, *98*, 1558–1565.
177. Gordon, D.; Chen, R.; Ho, J.; Coote, M.L.; Chung, S.-H. Rigid body Brownian dynamics as a tool for studying ion channel blockers. *J. Phys. Chem. B* **2012**, *116*, 1933–1941.
178. Yu, K.; Fu, W.; Liu, H.; Luo, X.; Chen, K.X.; Ding, J.; Shen, J.; Jiang, H. Computational simulations of interactions of scorpion toxins with the voltage-gated potassium ion channel. *Biophys. J.* **2004**, *86*, 3542–3555.
179. Gilson, M.K.; Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
180. Rashid, M.H.; Mahdavi, S.; Kuyucak, S. Computational studies of marine toxins targeting ion channels. *Mar. Drugs* **2013**, *11*, 848–869.
181. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
182. Chen, R.; Chung, S.-H. Conserved functional surface of antimammalian scorpion β -toxins. *J. Phys. Chem. B* **2012**, *116*, 4796–4800.
183. Jarzynski, C. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 329–351.
184. Baştuğ, T.; Chen, P.-C.; Patra, S.M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.
185. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; *et al.* Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
186. Yi, H.; Qiu, S.; Wu, Y.; Li, W.; Wang, B. Differential molecular information of maurotoxin peptide recognizing IK(Ca) and Kv1.2 channels explored by computational simulation. *BMC Struct. Biol.* **2011**, *11*, 3.
187. Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–6606.