CHAPTER 1

Seeing the Woods for the Trees: Understanding Venom Evolution as a Guide for Biodiscovery

BRYAN G. FRY*[a,b], IVAN KOLUDAROV[a], TIMOTHY N. W. JACKSON[a,b], MANDË HOLFORD[c], YVES TERRAT[d], NICHOLAS R. CASEWELL[e], EIVIND A. B. UNDHEIM[a,b], IRINA VETTER[b], SYED A. ALI[a,b,f], DOLYCE H. W. LOW[a], AND KARTIK SUNAGAR[g,h]

[a]Venom Evolution Lab, School of Biological Sciences, The University of Queensland, St. Lucia, Queensland 4072, Australia; [b]Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Queensland 4072, Australia; [c]The City University of New York-Hunter College & CUNY Graduate Center and The American Museum of Natural History, New York, New York, USA; [d]Montréal University, Research Institute in Plant Biology, Montreal Botanical Garden, Montreal, Québec, Canada; [e]Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK; [f]HEJ Research Institute of Chemistry, International Center for Chemical and Biological Sciences (ICCBS), University of Karachi, Karachi 75270, Pakistan; [g]CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal; [h]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
*E-mail: bgfry@uq.edu.au

RSC Drug Discovery Series No. 42
Venoms to Drugs: Venom as a Source for the Development of Human Therapeutics
Edited by Glenn F. King

Published by the Royal Society of Chemistry, www.rsc.org

1.1 The Fundamental Problems

The majority of commercial drugs being used today in both developed and developing countries are based on natural products.[1] Most of these products are based upon plants, but research into animal venoms holds great potential for the discovery of novel medicinally useful natural products.[2,3] Knowledge of the evolutionary origins of venom proteins/peptides and the forces shaping the biodiversity seen today is crucial for efficient biodiscovery. In addition, efficient utilisation of venom toxins in drug design and development cannot be achieved without recognition of the true biochemical, ecological, morphological, and pharmacological diversity of venoms and associated venom systems. A major limitation of the use of venom proteins thus far has been the very narrow taxonomical range studied. Entire groups of venomous animals remain virtually ignored. Those that have been examined have apparently been selected due to their medical significance or ease of collection, rather than as a result of their ecological or evolutionary uniqueness.

Venom is defined as "a secretion, produced in a specialised gland in one animal and delivered to a target animal through the infliction of a wound (regardless of how tiny it may be), which contains molecules that disrupt normal physiological or biochemical processes in the victim so as to facilitate feeding or defence by the producing animal".[4] This definition encompasses creatures normally considered venomous (*e.g.*, scorpions, snakes, and spiders) as well as animals that have not been traditionally recognised as such (*e.g.*, leeches, ticks, and vampire bats). Acknowledgement of the evolutionary analogy of the recruitment and use of toxins in all these animals increases the number of known independent occasions in which venom has evolved independently. In addition, this acknowledgement improves our understanding of the factors underlying the evolution of venoms and their associated proteins while also drawing attention to the vast pool of unstudied toxins. Venom has been a key innovation in the evolutionary history of an incredibly diverse range of animals. Even using the traditional definition of venom, venom systems are believed to have evolved independently on at least 20 occasions in extant lineages (Figure 1.1). Intriguing fossil evidence has also led to speculation about the possibility of extinct venomous lineages represented by the theropod dinosaur *Sinornithosaurus*[5] and the extinct pantolestid mammal *Bisonalveus browni*.[5] If lineages such as ticks, leeches, vampire bats, *etc.* are rightfully recognised as venomous, the number of independent evolutionary events in which venom has arisen increases to over 30.

The evolutionary selection pressure upon defensive venoms (*e.g.*, those of fish and bees) is largely directed at the development of streamlined venom that has the primary action of immediate, intense localised pain.[6–8] In contrast, predatory venoms are shaped by a classic co-evolutionary arms race, where evolving venom resistance in prey and the evolution of novel venom composition exerts reciprocal selective pressures on one another in a situation that conforms to the Red Queen hypothesis of Van Valen.[9] Powerful

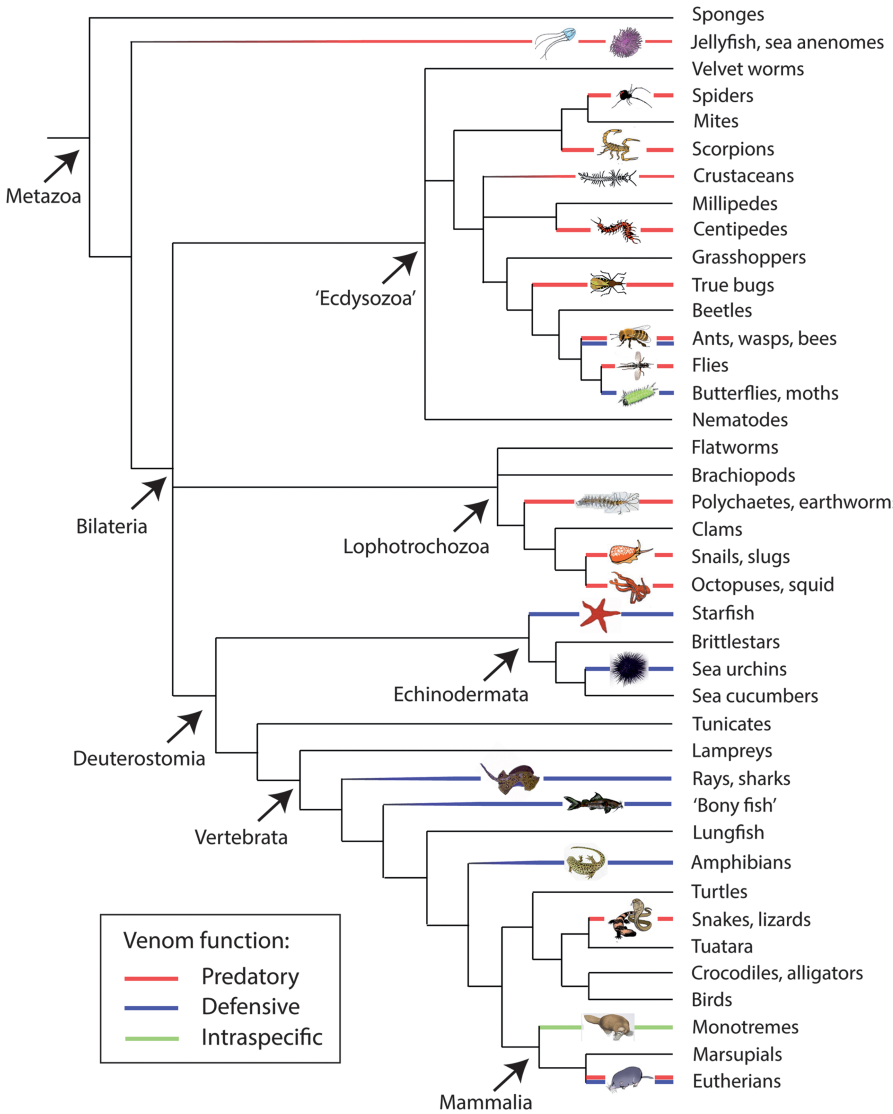


Figure 1.1 Schematic tree of venomous life in the animal kingdom. Coloured branches indicate lineages that include members with venom systems. Phylogeny based on the tree of life presented in Pennisi.[145] Note that a number of animal lineages have been pruned from the tree. Adapted from ref. 34.

purifying selection pressures acting on predatory venoms for millions of years have resulted in highly complex modern venom arsenals that consist of potent compounds with exquisite target specificity. Variation in venom composition is not only observed between different lineages, but also between the closely related species within a clade.[10] Intraclade differences in

venom composition often arise as a result of the evolution of prey-specific toxins in species with specialised diets.[11–13] Significant variation in venom profile has even been demonstrated within individual species with widespread geographical distributions.[14,15] Venom can also vary intraspecifically as the result of numerous other factors, including sibling differences[16] and ontogenic changes in prey preference[17] or behaviour. In Sydney funnel-web spiders (*Atrax robustus*), juvenile male spiders and female spiders of all ages have similar insecticidal predatory venoms, whilst sexually mature males (who stop feeding and leave the burrows in search of females) have a vertebrate-specific defensive venom.[18] It is this adaptive complexity and innovation that makes predatory venoms ideal candidates for the discovery of therapeutic lead compounds.

The majority of venom components have evolved to target physiological systems reachable by the bloodstream. In particular, the neurological and haemostatic systems have been convergently targeted *via* a myriad of innovative pathways (Figure 1.2).[4] A consistent feature of venom proteins is a stable molecular scaffold of cross-linked cysteines[19] (see Chapter 2 for further details of disulfide-rich toxin scaffolds); this characteristic appears to facilitate modification of non-structural residues, which in turn facilitates protein neo-/sub-functionalisation. A remarkable degree of convergence exists not only in terms of toxin molecular scaffolding, but also in target specificity and bioactivity.[4] The superimposition of sequences from functionally convergent toxins reveals tremendously useful information regarding structure–function relationships. An example of this is the platelet-aggregation inhibiting RGD tripeptide motif. This motif has been independently derived on numerous occasions within a myriad of distinct protein scaffolds, ranging from snakes (two different occasions: disintegrins and three-finger toxins) to a wide variety of invertebrate species, including ticks (*e.g.*, *Ixodes* spp., *Argas* spp., *Rhipicephalus* spp., *Amblyomma* spp.), tabanid flies (*e.g.*, *Tabanus* spp.), true bugs (*e.g.*, *Triatoma* spp., *Rhodnius prolixus*), mosquitoes (*e.g.*, *Anopheles* spp., *Aedes* spp., *Culex* spp.), sand flies (*e.g.*, *Lutzomyia* spp., *Phlebotomus* spp.), leeches (*e.g.*, *Macrobdella* spp., *Placobdella* spp.), and worms (*e.g.*, *Ancylostoma* spp.).[4,20] This reinforces the fact that biological targets within prey animals are the primary drivers of the evolution of toxin structures.

Snakes, spiders, scorpions, marine cone snails, and sea anemones represent the majority of venomous organisms that have been studied, with other venomous lineages remaining neglected. Moreover, even within these well-studied lineages, there has been a significant taxonomical bias. Partly as a result of the limited taxonomic range studied, the majority of known venom components remain poorly understood, and it is likely that many more venom components await discovery. The complex nature of venom makes it energetically expensive to produce. Hence, most venomous organisms have evolved a highly sophisticated cocktail that can efficiently aid in predation and/or defence, even when secreted in very small quantities. The small amount of venom produced by many venomous organisms was a major obstacle that impeded venom exploration in the past. Even in snakes, which

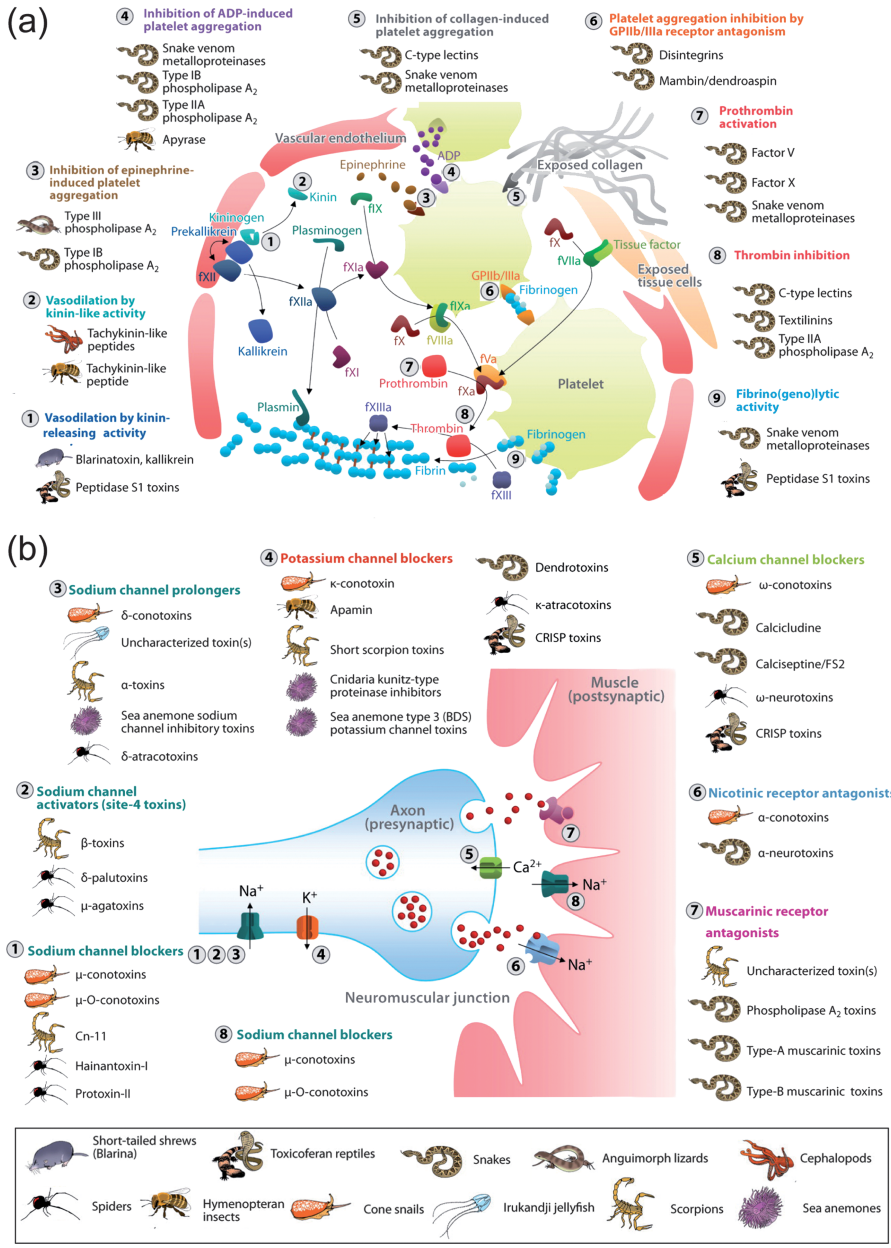


Figure 1.2 Convergence of toxin action in the animal kingdom. Sites of haemotoxic (A) and neurotoxic (B) convergence in animal venoms. Adapted from ref. 4.

may produce copious amounts of venom, particularly large amounts (multigram) were necessary for the discovery novel venom components that are secreted in miniscule amounts. For smaller animals such as spiders, the tiny amounts secreted made many species impossible to investigate using protein-based approaches.

Another impediment to venom exploration has been the difficulty of extracting venom from species that do not store secreted venom in readiness for delivery, or that have venom delivery systems that are difficult to access or stimulate. For instance, the venom delivery apparatus of non-front-fanged snakes is located at the back of the mouth and, unlike those of many front-fanged snakes, the venom glands do not contain an appreciable lumen for venom storage, instead only secreting venom as required. Hence, it was very difficult and time-consuming to obtain sufficient quantities of venom from such snakes for the "proteome-only" oriented venom research of the past. Even chemical stimulation of venom secretion (*e.g.*, injection of pilocarpine into the venom gland) has been unsuccessful in overcoming some of the aforementioned complications, impeding venom research in a large group of other organisms (*e.g.*, coleoids, centipedes, non-front-fanged snakes, spiders, and vampire bats). In some cases, these difficulties have been overcome through the application of considerable amounts of time and effort. For example, it took over 200 pilocarpine-stimulated milkings (venom extraction) of *Coelognathus radiatus* to obtain 110 mg of crude venom, which yielded 10 mg of pure α-colubritoxin in the first study of a three-finger toxin (3FTx) from a non-elapid snake.[21] In other cases, however, the difficulties described above have proven impossible to surmount.

With the advent of next-generation RNA sequencing, venom exploration has become more efficient, as researchers can now rapidly construct transcriptome libraries of entire venom glands, without depending on proteinaceous venomous secretions directly. While transcriptomics will rapidly yield full-length precursor sequences, the prediction of propeptide cleavage sites, other than those that are conventionally dibasic, may be impossible.[22] Moreover, transcriptomics alone cannot unravel post-translational modifications (PTMs), which are often crucial for the biological activities of venom components. For example, a sulfo-tyrosine PTM is required for the bioactivity of lizard venom cholecystotoxin; synthetic analogues lacking this PTM are completely inactive.[23] For these reasons, a combined proteomic–transcriptomic approach is essential for the most effective venom exploration (see Chapter 3).

Genome information remains scarce for species other than those routinely used as model organisms in genetics research. To date, genomes are available for only a few venomous animals. Knowledge of the location and organisation of venom-encoding genes can greatly increase our understanding of their molecular evolutionary history. Availability of genomic information will facilitate easier amplification of specific venom-encoding genes. Moreover, this will enable researchers to use small amounts of tissue or other non-destructive samples for sequencing of venom-encoding genes, not only

making venom exploration in rare venomous organisms easier and more sustainable, but also overcoming the difficulties of obtaining permits for destructive sampling for research.

Another major problem that has affected venom research is the difficulty of obtaining the venomous animals themselves. Most well-studied venomous organisms have been those that are locally common in the regions in which the research was performed. Researchers often restrict their venom collection to species represented in local serpentariums or that are available from other venom suppliers. Samples acquired from third parties in this manner are often associated with uncertainties regarding geographical origin and sometimes even basic taxonomy. For example, the Sigma pharmaceutical catalogue entry for *Oxyuranus scutellatus* venom (see http://www.sigmaaldrich.com/catalog/product/SIGMA/V3129) states "This venom may be from subspecies *O. s canni* (Papuan taipan) or *O. s scutellatus* (Australian taipan) or a mixture from both"; with a note that "Physical characteristics are almost identical". This is despite the existence of abundant research showing that venoms may vary appreciably across a relatively short continuous geographical range, let alone the sort of variance that may occur between completely disjunct localities. Considerable differences in toxicity and antivenom coverage have recently been demonstrated for *O. s. canni* and *O. s. scutellatus*,[24] which highlights the fact that disregarding the geographical origin of samples is unacceptable in venom research.

The taxonomical bias in toxinology is starkly evident when sequenced toxins are mapped against organismal diversity. For example, in elapid snake venom research, two genera (*Bungarus* and *Naja*) account for almost 40% of all published sequences (Table 1.1). Moreover, almost 40% of all 3FTxs have been sequenced from *Naja* alone. Despite the diversity of toxin forms present, some toxin types are known from transcriptomic studies only. Similarly, of the 3FTxs known from the non-front fanged snake lineages, the majority are known from a single transcriptomic study.[25] Only three studies have characterised the bioactivity of fully-sequenced 3FTx from non-front-fanged snake venoms.[26–28] This bias is not unique to snakes, as the other venomous lineages that have received toxinological attention have suffered similar levels of taxonomical bias. For example, although scorpion venoms have received more research attention than the venoms of any other lineage, only 50 or so of the approximately 1700 species of scorpion recognised today have been examined. The major focus has been on basal families such as Buthidae, which account for more than 50% of all known scorpion toxin sequences. These basal families are known to have separated from all other scorpion families about 350 million years ago,[29,30] suggesting that there is likely to be a plethora of novel venom components that remain undiscovered in the other families. Similarly, despite spiders being the most speciose group of venomous animals, represented by ~45 000 recognised species, venom exploration remains primarily restricted to large mygalomorph species. For example, tarantulas account for more than one quarter of all spider toxins isolated to date, although they represent only ~2% of the

Table 1.1 Toxins sequenced from elapid venoms[a]

Genus	*Total*	*3FTx*	*AVIT*	*Acn*	*CRiSP*	*fV*	*fX*	*Kun*	*Lec*	*LAO*	*NP*	*NGF*	*PLA_2*	*SP*	*SVMP*	*Ves*	*Wap*
Acanthophis	**13**	4									1		8				
Acalyptophis	**2**	2															
Aipysurus	**13**												13				
Aspidelaps	**5**	3			1								1				
Astrotia	**4**	4															
Austrelaps	**63**	19			3			6					32				3
Boulengerina	**5**	3													2		
Bungarus	**218**	80		1	1			39	5	2	1	1	87	1			
Calliophis	**3**	1											2				
Cryptophis	**20**				1		1	6	3		2	1	4				2
Demansia	**40**	3			1		2	6	3	1		3	17				4
Dendroaspis	**59**	47	1					9			2						
Drysdalia	**38**	25			1			9					3				
Enhydrina	**3**	2											1				
Hemachatus	**11**	8						2					1				
Hoplocephalus	**16**	1			1		1	3			1		6		1	1	1
Hydrophis	**10**	9											1				
Laticauda	**60**	24			1								35				
Lapemis	**9**	6			2									1			
Micropechis	**2**												1		1		
Micrurus	**74**	44						2	5		3		20				
Naja	**310**	226			11			8	4	5	2	2	37	2	12	1	
Notechis	**44**	2			1		3	4	5	1	2	3	21				2
Ophiophagus	**57**	44			1			2	2	1	1		3	1	1	1	
Oxyuranus	**126**	15			5	2	2	12	9	3	12	3	60				3
Pelamis	**1**	1															
Pseudechis	**85**	3			2		1	13	12	1	6	4	37				6
Pseudonaja	**47**	12			1	2	1	7	5			2	15		1		1
Tropidechis	**20**	3			2		1	2	3			0	7				2
Walterinnesia	**9**	3						4					2				
Total	**1367**	**594**	**1**	**1**	**35**	**4**	**12**	**134**	**56**	**14**	**33**	**19**	**414**	**5**	**18**	**3**	**24**

[a]3FTx = three-finger toxin, Acn = acetylcholinesterase, fV = factor V, fX = factor X, Kun = kunitz, Lec = lectin, LAO = L-amino oxidase, NP = natriuretic peptide, NGF = nerve growth factor, PLA_2 = phospholipase A_2, SP = serine protease, SVMP = snake venom metalloprotease, Ves = vespryn, Wap = waprin

taxonomic diversity of spiders.[31] Furthermore, it is suggested that the currently recognised species constitute only ~25% of existing species.[32] Spiders evolved from the stem arachnid ancestor about 300 million years ago during the Carboniferous period. Spider venoms contain a range of low molecular weight peptides and proteins that are neurotoxic, haemotoxic or cytotoxic in activity. It is likely that the highly complex nature of the venom is responsible for the tremendous success and diversification of the spiders as a group. Continuing this theme, the venom has not been thoroughly characterised from a single species of centipede amongst the 3300 species of centipedes known today. The forcipules, or poison claws, which are modified front legs used for delivering venom into the prey, have been identified in centipede fossils dating back to the early Devonian period, 400 million years ago. This suggests that centipedes, along with scorpions, possess one of the most ancient venom delivery apparatuses.[33] Despite this, centipede venom research is very much in its infancy.

1.2 The Solutions

The most efficient venom exploration approach is multidisciplinary and encompasses various fields and techniques, including:

- Organismal selection based upon phylogenetic position and ecological niche occupied
- Transcriptomics and *in silico* studies
- Molecular evolution and phylogenetics of toxins
- Proteomics
- Bioactivity testing

1.2.1 Taxon Selection

In order to maximise the efficiency of biodiscovery efforts, researchers should endeavour to examine the most diverse range of taxa possible. As highlighted elsewhere in this chapter, toxinological research in the past has focussed on species that were easy to acquire venom from or that were the most "medically significant", where medical significance is defined as "the danger posed to a human by a bite/sting". This reliance on common species, or the tiny minority of venomous species that are dangerous to humans, has resulted in the fact that the vast majority of venomous animals remain under-researched or completely ignored.

There are two primary drivers of diversity in venom evolution, both of which should be taken into account. The first is phylogenetic distance—species that are distantly related are likely to have more divergent venom components than species that are closely related. For this reason, researchers can widen their net in the search for novel compounds by examining members of as many genera as possible within their chosen clade, rather than concentrating solely on one genus.

The second primary driver of venom diversity is ecological distance, specifically differences in prey type, foraging mode and prey-handling behaviour. The majority of venomous animals use their toxic secretions to aid in prey subjugation, and hence selection pressures driving the evolution of venom components originate, in large part, from the prey species.[34] It is likely that different venom components are more suited to subduing different types of potential prey animals, therefore venomous species, even those that are closely related, that feed on divergent prey types are likely to have divergent venom compositions. The foraging mode largely determines the type of prey available to a venomous animal, but it also determines prey condition. Prey conditions, such as body temperature, activity level, *etc.* may affect the physiological action of toxins; therefore venomous predators that feed on the same prey types, but feed when the prey is in a different condition, may have divergent venom compositions. For example, snakes that forage at night for sleeping diurnal lizards will encounter inactive prey with a low body temperature and may therefore have less need to quickly disable prey that are too cold to resist effectively but may also rely on toxins that have a temperature-independent activity. Snakes that forage for the same lizard species during the day will encounter active prey with high body temperatures and may therefore rely on toxins that quickly disable prey to prevent it escaping or injuring the snake, and may be able to utilise toxin types that have a temperature-dependent activity. Similarly, prey-handling behaviour may influence venom composition. Venomous predators that strike and release their prey may rely on toxins that rapidly disable motor functions in order to ensure that prey does not travel far after envenomation and is therefore easy to track down. Venomous predators that strike and hold onto prey may have less need to disable motor functions and may also have the option of physically subduing prey; that is, they may use a combined physical and chemical attack (*e.g.*, use of powerful pincers and venom in scorpions, the use of constriction and venom in snakes, or harpoon-shaped radulars filled with venom in marine snails). Good examples of the fact that investigation of atypical subjects of toxinological research yields "low-hanging fruit" for biodiscovery include recent studies on organisms as diverse as Antarctic octopods[35] and lizards previously considered "non-venomous".[23,36]

Although venom is ubiquitous within the Octopoda, the majority of toxinological work on this order has been focussed on the medically significant *Hapalochlaena* sp. (blue-lined octopods) that harbour tetrodotoxin (TTX) produced by endosymbiotic bacteria.[37] Prior to 2010, nothing was known about the composition of the venom of octopods from the waters of the Southern Ocean in the vicinity of Antarctica. Inhabiting waters with sub-zero temperatures has placed extreme selection pressures on the venoms of these octopods and not only were two new toxin classes not previously known from octopods identified in their venoms, but the activity of enzymatic toxins contained therein displayed extreme cold-adaption, with enzymes generally more active at 0 °C than 37 °C.[35]

Although it was previously believed that venom was restricted within lizards to the two species in the family Helodermatidae, the discovery that toxin-secreting oral glands are an ancestral trait of toxicoferan squamate reptiles[38,39] vastly increased the number of lizard species of potential interest to toxinological researchers. Subsequent investigation of the venom systems of these lizards has not only increased our understanding of the evolution of venom in squamate reptiles, but has also resulted in the discovery of a number of new peptide types, including three that affect the cardiovascular system and are thus of potential interest in drug design and development.[40]

As investigators continue to widen the scope of their toxinological research, it is inevitable that many new toxins of interest from a biodiscovery perspective will be uncovered. Understanding the evolutionary pathways that venom travels and the evolution and ecology of the whole organism in which venom evolves serves as an effective guide for biodiscoverers. Through this understanding they may gain valuable knowledge of the most efficient directions in which to cast their nets. Indeed, viewing the study of venom evolution and the search for novel toxins as separate disciplines has become an outmoded and unconstructive viewpoint. These areas of research are two sides of the same coin and complement one another to the extent that a discovery in the field of venom evolution will almost invariably uncover a valuable new resource for biodiscovery.

1.2.2 Transcriptomics

With the advent of high-throughput sequencing technologies, scientists can generate datasets of considerable size. These advances have had considerable benefit for the field of toxinology. Recent studies have demonstrated the effectiveness of applying these technologies to the tasks of identifying novel toxin components and exploring diverse venom-encoding gene families that have previously been characterised.[37,39–46] Several methods of toxin annotation have been employed for proteomic and transcriptomic datasets, mainly based on "BLAST gold standards".[47–51] Complementary phylogenetic analyses of toxin gene families have also revolutionised our understanding of the origins and evolution of toxins. However, phylogenetic analyses of toxin gene datasets have been hampered by significant problems: (i) several toxin types have very short sequences and consequently few informative phylogenetic positions are identified on alignments (*e.g.*, conopeptides from cone snails, crotamines in rattlesnakes); (ii) it has been shown that signal, pro and mature peptides may adopt different evolutionary pathways, leading to low statistical support among phylogenies; (iii) some toxin gene families experience tremendous diversifying selection pressures,[52] resulting in a lack of phylogenetic resolution as each amino acid site experiences multiple changes, which poses a major problem for multiple sequence alignments; (iv) performing phylogenetic analyses on large datasets can be problematic if strong statistical support is needed, especially for maximum likelihood and Bayesian inferences.

One way of addressing these issues is to employ genetic-network-based approaches to unravel the evolutionary histories of toxins. Genetic networks have been extensively used to study genes that follow reticulate evolutionary pathways, such as those that experience regular recombination and lateral gene transfers.[53] A genetic network (or a graph) is a mathematical model of pairwise relations among entities named nodes. The nodes in the network are linked by edges representing the connections or interactions between these entities. Building a network allows nodes to be grouped into connected components. Nodes and interactions can be of diverse nature (directed or undirected interactions, metabolic pathways, transcriptional regulation processes, *etc.*). In the case of toxin gene networks, the nodes represent amino acid or nucleotide sequences of toxins and the connections are represented as "% similarity" shared between two sequences. A new software called Evolutionary Gene & Genome Network (EGN) has recently been used for metagenomic analysis[54] and is now available to the public.[146] This software, which is written in PERL, has been used for building toxin gene networks by making use of the manually curated ToxProtDb database.[55] It begins with the comparison of all sequences from a given dataset with one another using sequence similarity tools such as BLAST or BLAT, which are implemented in the software itself. The tabular output generated by the software contains parameters such as percent ID, percent similarity, e-value and score, which are then processed by EGN to generate graphs. The graph can be further visualised using software such as Cytoscape[56] and Gephi.[57] The primary benefit of these analyses is their ability to process datasets generated with high-throughput techniques.

In order to discuss some of the main advantages of network-based analyses, we will describe a few examples.

The ToxProtDb dataset includes 5361 manually reviewed toxin sequences. We first excluded sequences that were less than 40 residues in length. Out of the remaining 4339 sequences, we were able to cluster 3071 sequences into 93 toxin groups or connected components (Figure 1.3A) using a BLAST similarity search (minimal e-value: 1E-05; minimal hit identity threshold: 20%). Using the aforementioned parameters, we were unable to find any matching sequences for the remaining 1268 sequences, which were therefore considered to be "singletons". Nodes were coloured according to the taxonomic classification of sequences. Graphical representation of the connected components not only reveals the diversity that exists among these examined groups (Figure 1.3B), but also the major convergent recruitment events of different toxins (Figure 1.3C and D). In our analyses, Kunitz proteases came out as one of the convergently recruited toxins in spiders, cone snails, insects, and snakes. The discovery of the similarity between U_{13}-theraphotoxin-Cj1a isolated from the spider *Chilobrachys jingzhao* and M-conopeptide isolated from *Conus vexillum,* however, was more surprising. As illustrated in Figure 1.3A, this similarity is not restricted to cysteine pattern alone. To our knowledge, this level of convergence has not been documented before, and its discovery reinforces the value of network-based analyses.

Network-based analyses can also be focussed on particular taxa. Figure 1.3D represents the evolutionary network analysis of snake venom proteins. Of 1701 sequences available on ToxProtDb, 1351 sequences form 24 different connected components (e-value threshold: 1E-05). This network perfectly illustrates the predominance of phospholipases and 3FTxs in snake venoms. The shape of each connected component also shows that some families such as the metalloproteinases are highly diverse, while others such as phospholipases are highly conserved.

Network-based analyses can help in processing large datasets quickly. Further, the network analyses offer numerous statistical frameworks for comparative purposes. Genetic diversity among each connected component could be evaluated using simple statistics such as the clustering coefficient, the sum of edges divided by total number of possible edges, or network diameter. Highly sophisticated parameters concerning whole network shape or individual nodes could also be estimated (degree distributions, neighbourhood connectivity, between-ness centrality, *etc.*) using the Cytoscape and Gephi plugins. A basic representation of comparative diversity is illustrated in Figure 1.3C. The most compact connected component includes highly similar sequences as shown by the Weblogo consensus.[58] A more diverse connected component is represented as well, with the sequences exhibiting a low level of similarity outside cysteine residues.

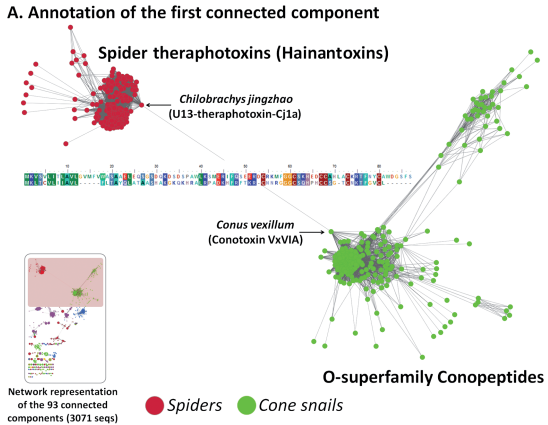
To summarise, network-based approaches offer several advantages when applied to toxin datasets. They: (i) allow efficient annotation of toxins in large datasets; (ii) efficiently process comparative analyses using previously identified toxins; (iii) help in identification of closest relatives for further phylogenetic analysis; (iv) significantly improve multiple sequence alignments and clustering; and (v) facilitate visualisation of reticulate evolutionary pathways that cannot be represented using bifurcating phylogenetic trees (nodes could have numerous connections instead of the usual tree nodes with only one parent).

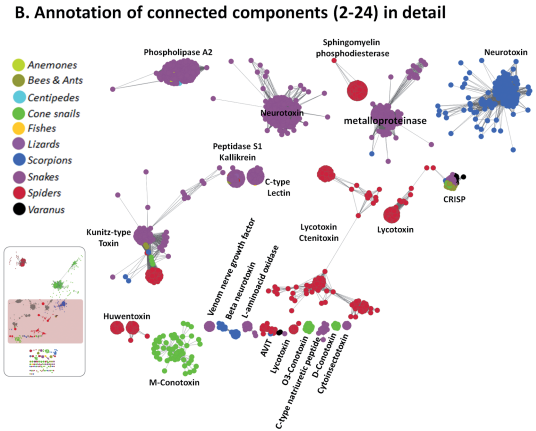
1.2.3 Sequence Retrieval and Alignment

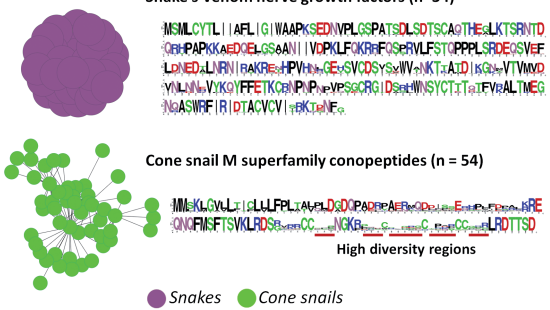
Newly sequenced toxins may be identified by comparison with previously characterised sequences using search programs such as BLAST[59] implemented in various public sequence databases like the UniProtKB protein database (http://web.expasy.org/blast) and NCBI (http:/ncbi.nlm.nih.gov/). This is often employed to retrieve additional homologous sequences, already deposited in the aforementioned databases, so as to increase the number of sequences in the dataset. Resultant sequence sets are aligned using programs such as CLC Main Workbench, CLUSTAL,[60] MUSCLE,[61] *etc.*

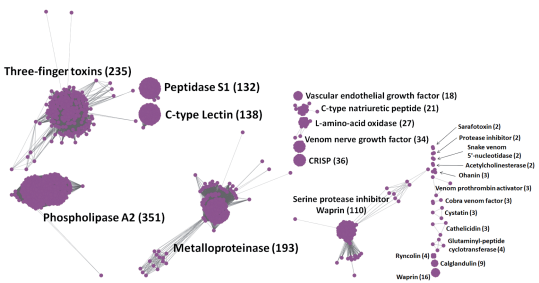
1.2.4 Phylogenetic Analyses

Phylogenetic analyses are performed to reconstruct the molecular evolutionary history of each toxin type. Datasets are analysed with methods such as Bayesian inference implemented in MrBayes[62] or maximum-likelihood


A. Annotation of the first connected component
Spider theraphotoxins (Hainantoxins)
Chilobrachys jingzhao
(U13-theraphotoxin-Cj1a)
Conus vexillum
(Conotoxin VxVIA)
O-superfamily Conopeptides
Network representation
of the 93 connected
components (3071 seqs)
Spiders
Cone snails


B. Annotation of connected components (2-24) in detail
Anemones
Bees & Ants
Centipedes
Cone snails
Fishes
Lizards
Scorpions
Snakes
Spiders
Varanus
Phospholipase A2
Neurotoxin
Sphingomyelin
phosphodiesterase
metalloproteinase
Neurotoxin
Peptidase S1
Kallikrein
C-type
Lectin
CRISP
Kunitz-type
Toxin
Lycotoxin
Ctenitoxin
Lycotoxin
Huwentoxin
M-Conotoxin
Venom nerve growth factor
Beta neurotoxin
L-aminoacid oxidase
AVIT
Lycotoxin
O1-Conotoxin
C-type natriuretic peptide
D-Conotoxin
Cysteine protease


C. Comparison of compact versus scattered connected components
Snake's Venom nerve growth factors (n=34)
Cone snail M superfamily conopeptides (n = 54)
High diversity regions
Snakes
Cone snails


D. Evolutionary network analysis of snake's venom proteins
Three-finger toxins (235)
Peptidase S1 (132)
C-type Lectin (138)
Phospholipase A2 (351)
Metalloproteinase (193)
Vascular endothelial growth factor (18)
C-type natriuretic peptide (21)
L-amino-acid oxidase (27)
Venom nerve growth factor (34)
CRISP (36)
Serine protease inhibitor
Waprin (110)
Sarafotoxin (2)
Protease inhibitor (2)
Snake venom
5'-nucleotidase (2)
Acetylcholinesterase (2)
Ohanin (3)
Venom prothrombin activator (3)
Cobra venom factor (3)
Cystatin (3)
Cathelicidin (3)
Glutaminyl-peptide
cyclotransferase (4)
Ryncolin (4)
Calglandulin (9)
Waprin (16)

implemented in various applications such as PhyML,[63] Randomized Axelerated Maximum Likelihood (RAxML),[64] *etc.* Although both the aforementioned methods can generate accurate phylogenetic trees, the Bayesian methods are widely preferred.

1.2.5 Test for Recombination

Recombination is a process in which two different molecules of DNA, usually homologous, exchange genetic material with one another. Recombination can mislead phylogenetic interpretations.[65] Hence it becomes essential to identify and eliminate recombinant sequences from datasets before conducting phylogenetic analyses. All of the contemporary algorithms that are designed to detect selection pressures assume the absence of recombination in the dataset. Moreover, the variations that result from recombination events may resemble those that result from adaptive evolution,[66] and the process also leads to apparent substitution rate heterogeneity.[67] Thus, recombination can have an impact on evolutionary selection analyses as well. One could either remove identified recombinant sequences from datasets altogether or allow the recombinant sequences to have independent phylogenetic histories while estimating selection pressures. The former can be done by employing various programs such as SIMPLOT,[68] RDP,[69] TOPALi,[70] *etc.* while the latter is achieved through methods such as Single Breakpoint algorithm (SBP) and Genetic Algorithm for Recombination Detection (GARD), implemented in the HyPhy package[71–73] Potential breakpoints can be

Figure 1.3 Network analysis of toxin genes. (A) The convergence between O-superfamily *Conus* venom peptide and typical spider-venom peptides. These two large groups of toxins (based on ToxProtDb data) are connected through the toxin U_{13}-theraphotoxin-Cj1a from the tarantula *Chilobrachys jingzhao* and conopeptide VxVIA from the venom of *Conus vexillum*. They share an identical cysteine framework (C–CC–C–C) and similar residues in the mature toxin sequence. (B) The convergent recruitment of various toxin groups. Some groups are exclusively found in a given taxonomic group (*e.g.*, huwentoxins in spiders) while highly similar cysteine-rich secretory proteins (CriSPs) are found in various taxa such as cone snails, bees, scorpions, spiders, and lizards. (C) Demonstration that gene network analysis also facilitates the characterisation of genetic diversity for a given family. The shape of the network shows the intrinsic diversity within each group. The M-conotoxin group from cone snails includes sequences with a low degree of conservation outside the cysteine framework, giving rise to a dispersed network shape, whereas snake-venom nerve growth factors are highly similar, giving rise to a compact shape. (D) The composition and the diversity of snake venom protein families. PLA_2s and three-finger toxins are the largest groups of snake venom proteins studied so far. This may reflect the composition of snake venom and/or a biased interest of scientists for certain types of toxins.

detected using the small sample Akaike information Criterion (cAIC) and sequences can be compartmentalised before conducting phylogenetic or evolutionary analyses.

1.2.6 Identifying Evolutionary Selection Pressures

It is essential to recognise the evolutionary selection pressures that shape venom components, including those that act on different regions of the same venom component. This not only reveals the functional importance of different toxin domains but also provides a wealth of information for drug design and development. Many essential genes that exist in single copy within the genome evolve through the regime of negative selection, whereby a single non-synonymous mutation in the functional domain can lead to the death of the organism (lethal mutations), resulting in the elimination of such mutations from the population. By contrast, the genes that encode immunoglobulins are required to incorporate variation to combat the diversity of pathogens the body encounters. Similarly, predatory venomous organisms benefit from producing a variety of toxin forms to stay ahead in the chemical arms race with their prey. Thus, negative selection eliminates mutations that have harmful effects on the fitness of the organism, while positive selection generates mutations/variations that increase the fitness of the organism. Parallel to these two mechanisms, neutral evolution causes non-functional synonymous and non-synonymous mutations to accumulate in equal proportions in neutral alleles (those that do not affect the fitness of the organism). It is worth noting that the neutral theory of molecular evolution considers neutral mutations to be far more common, and thus have more impact on sequence composition, than positive selection.[74]

Venomous predators stay ahead in the chemical arms race with their prey as a result of the rapidity with which toxins accumulate variations. Formation of toxin multigene families is one of the primary ways in which venom-encoding genes diversify. Recurrent duplication events result in the formation of new copies of venom-encoding genes, which evade pre-existing negative selection pressures. Following duplication, they are reinforced through positive selection and may be neofunctionalised to produce a myriad of different peptides with novel biochemical properties. Identifying regions that accumulate variation is not only important for understanding the evolutionary history of the toxin, but is also useful for novel methods of antivenom production. It is logical to target regions that are under the constraints of negative selection, and hence are least likely change over short periods of time, rather than targeting regions that accumulate tremendous variation and are thus less likely to cross-react with antibodies raised against other isoforms, limiting the usefulness of antivenoms based on these epitopes.

Historically, positive selection has been detected as a ratio (ω) of non-synonymous (Ka or dN or α) to synonymous (Ks or dS or β) substitutions. Synonymous substitutions are considered neutral in terms of selection, as they do not change the amino acid sequence, while non-synonymous

substitutions are considered a function of selective pressure on the protein since they change the primary structure and may affect function as well. Negative selection pressure will reduce the accumulation of deleterious non-synonymous mutations, effectively reducing the ω value (dN/dS) to less than 1. By contrast, positive selection will increase non-synonymous mutations relative to synonymous mutations, increasing the ω value to more than 1. If the protein evolves neutrally, it will accumulate synonymous and non-synonymous mutations in equal proportions, resulting in an ω value of 1. Sophisticated likelihood models of coding-sequence evolution,[75,76] implemented in CODEML of the PAML[77] package, have been popularly utilised to evaluate selection pressures. Lineage-specific models can be utilised to identify selection pressures acting across lineages in a phylogenetic tree. The two-ratio model is often employed for identifying selection pressures across different lineages. However, this model requires the branches evolving under positive selection to be defined *a priori*, which is often not possible. Moreover, the lineage-specific models fail to identify regions in proteins that might be affected by episodic selection pressures, and hence they can underestimate the strength of selection. Hence, site-specific models are employed that estimate positive selection statistically as the non-synonymous-to-synonymous nucleotide-substitution rate ratio (ω), where a ratio of significantly greater than 1 indicates positive selection. While lineage-specific models estimate sites under positive selection across lineages and site-specific models along sites, the branch-site models identify codon sites under selection across lineages and along sites, making them valuable tools for estimation of selection pressures. However, similar to the lineage-specific branch models (two-ratio model), the branch-site model also requires the *a priori* definition of lineages under selection.

Unlike the lineage-specific branch and branch-site models, the GA-Branch Test implemented in the HyPhy[78] package does not require the foreground and background branches to be defined *a priori*. The algorithm works on the principle that there could be many models that fit the data more closely than a single *a priori* hypothesis. Separating lineages into foreground (positively selected) and background (negatively or neutrally evolving) groups *a priori* leads to high rates of false positives and false negatives, especially when the mode of evolution of the background branches is different from the modelling assumptions. To overcome these limitations, a branch-site random effects likelihood (REL) model has been proposed, which estimates variations over sites and branches within the REL framework.[79]

1.2.7 Structural Analyses

Domains under selection can be depicted by mapping mutations over the structure of the protein as determined using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Because of the lack of structural information for most toxins, one may construct homology models using various servers such as the Swiss-model,[80–82] Phyre 2,[83] *etc.* Several

webservers are also available for automatically measuring selection pressures, employing various models of estimating selection, and furnishing homology models with mapped evolutionary conservations and variations.[84–86]

1.2.8 Proteomics

The search for novel toxins and other bioactive compounds in organisms other than those routinely used as model organisms in genetic research has traditionally been an onerous task involving activity-guided purification followed by partial sequencing from the N-terminus by Edman degradation. Full toxin sequences were either obtained by proteolytic digestion of pure proteins and sequencing of purified fragments or rapid amplification of cDNA ends (RACE) from venom gland RNA using primers designed from N-terminal partial protein sequences, followed by sequencing of the amplified product using Sanger sequencing. Although these methods of obtaining toxin sequences are time-consuming and in many ways "out-dated", the coupling of transcriptomic and proteomic techniques nevertheless remains one of the most powerful approaches for investigation of venoms and their components.

High-throughput proteomics is heavily reliant upon the existence of sequence templates, with the sequencing of entirely novel peptides and proteins still requiring substantial manual interpretation. Similarly, most bioinformatic pipelines for the processing and annotation of the ever growing next-generation sequencing (NGS) datasets still rely on sequence homology to tease out toxin sequences. While there are algorithms available for identification of putative toxin sequences without the use of homology searches, the physical presence of the predicted mature toxin in the venom cannot be confirmed without proteomic evidence. Furthermore the presence of PTMs, which can have a significant effect on activity and specificity of toxins,[25] is generally not reliably deduced from sequence information alone and must be confirmed by examination of the pure native toxin.

The ability to rapidly attain information on toxin sequences and PTMs is also useful beyond high-throughput description of venoms. For example, it is useful in identifying hits from activity screens, a process that could otherwise quickly become expensive and time consuming. By using transcriptomic data as a sequence template, the interpretation of top-down proteomic data can be largely automated and toxin identification, along with post-translational cleavage sites and any modifications, can generally be obtained within short time.

Thus, in isolation, both approaches are prone to becoming laborious when dealing with novel samples, but in disparate ways; one requiring a sequence template and the other requiring sequence validation. With the advent of NGS and the continuing development of faster, more sensitive mass spectrometers, a complementary approach allows for high-throughput verification of mature venom component sequences from transcriptomic data, including those of entirely novel toxins and any PTMs (see Chapter 3 for further details). The number of software suites (*e.g.*, Tandem, PEAKS, ProteinPilot) available

means that integration of the two datasets is largely automated, maintaining the high-throughput rate all the way through to data interpretation. Further investigation of sequences from the resulting venomic library can then be pursued on the basis of novelty, biochemical and structural properties, or homology to previously identified bioactive compounds.

1.2.9 Bioactivity Testing

Venoms have evolved over millions of years to target vital physiological processes within prey and predators with exquisite specificity and potency. The vast pharmacological cornucopia contained in venoms has: (i) contributed substantially to our understanding of human physiology and the pathophysiology of disease; (ii) provided important pharmacological modulators and tool compounds; and (iii) most importantly, led to the discovery of new drugs that have had significant positive impact on human health.

As we begin to appreciate the full complexity of venoms, with several hundreds to thousands of bioactive molecules contained in the venom of a single spider or cone snail,[87,88] understanding and delineating the biological activity of venoms and venom components becomes increasingly important. However, determining the molecular targets of venom components or isolating specific molecules with particular activity from venoms has proven challenging, as discussed in further detail in Chapter 4.

Early studies focussed on delineating the physiological and pharmacological effects of venoms resulting from *in vivo* administration to animals. Such studies were able to provide immediate insight into the physiological consequences of envenomation and led to the discovery of some of the most notable examples of drugs derived from venoms. For example, conotoxin GVIA was isolated following the observation that intracerebroventricular administration of venom to mice caused a shaking phenotype.[89–91] This was determined to occur as a result of inhibition of neuronal calcium channels, in particular $Ca_v2.2$,[92] and a related $Ca_v2.2$-selective conopeptide, MVIIA (ziconotide or Prialt®), was subsequently developed as a novel treatment for intractable pain[93,94] (see Chapter 9 for further details). Similarly, the observation that injection of venom from a Brazilian viper (*Bothrops jararaca*) caused a marked, sudden drop in blood pressure mediated through inhibition of angiotensin-converting enzyme (ACE), led to the development of modern antihypertensive ACE-inhibiting drugs (see Chapter 5 for further details).[95–97]

However, depending on the site of administration, such *in vivo* assays can produce false negative results. This was the case initially for several peptides from *Conus geographus* venom, which produced no effects after intraperitoneal injection in mice, but elicited profound behavioural changes after intracerebroventricular administration.[89] Similarly, *in vivo* studies are unlikely to detect biological activity that is not associated with overt behavioural or physiological changes. In addition, such studies may be no longer ethically acceptable, are time-consuming, low-throughput and clearly ill suited to large-scale screening of venoms for specific biological activity.

As an extension to *in vivo* studies, *ex vivo* (organ bath or isolated tissue preparation) experiments have been used for many years to characterise the physiological and pharmacological effects of venoms and venom components (see Chapter 4 for further details). *Ex vivo* preparations facilitate detailed pharmacological interrogation of the mechanisms of action underlying the biological activity of venom components. Exemplifying this strategy is the discovery of the χ and ρ classes of conotoxins from the venom of *Conus tulipa* and *Conus marmoreus*, which displayed inhibition and prolongation of electrically excited contractions, respectively, for ρ-TIA and χ-MrIA in a rat vas deferens tissue preparation[98] (see Chapter 9 for further details).

However, like *in vivo* studies, bioactivity testing in *ex vivo* tissue preparations is time-consuming and not compatible with high-throughput identification of venom components with specific action at selected therapeutic targets. Thus, in order to accelerate drug discovery efforts, high-throughput approaches are increasingly being applied to the isolation of bioactive components from venoms.

Taking advantage of the rich chemical diversity present in venom, high-throughput activity-guided fractionation has been used as an alternative strategy for the discovery and isolation of novel venom components. The aim of high-throughput screening (HTS) is to systematically isolate and characterise bioactive molecules from the complex mixture of peptides, proteins and small molecules present in venoms. This approach relies on the establishment of specific assays to examine the effect of novel compounds on a particular pharmacological target of interest, usually in cell lines heterologously expressing this target (see Chapter 4 for further details). As a minimum, it requires high assay sensitivity, accuracy, robustness, and reproducibility, often defined by a high Z′ score.[99] Assays developed for and amenable to HTS are as varied as the potential therapeutic targets themselves, and are able to assess activity of G-protein-coupled receptors, ligand- and voltage-gated ion channels, enzymes, or transporters, to name a few. Accordingly, the repertoire of HTS assays available now expands on more traditional approaches such as electrophysiological recordings and radioligand binding studies to include assays based on the detection of absorbance, bioluminescence, fluorescence, fluorescence polarisation, fluorescence-resonance energy transfer (FRET), and bioluminescence-resonance energy transfer (BRET) as well as label-free technologies and image-based high content screening (reviewed in ref. 100). The ability to miniaturise such assays, allowing the assessment of activity in 96-, 384- or 1536-well format, enables screening of activity from venom samples that are only available in very limited quantities, including venoms from cone snails, spiders, centipedes, and scorpions. In combination with transcriptomic approaches, thus circumventing the need for time-consuming and resource-intensive sequence analysis, HTS has the potential to vastly accelerate the discovery of novel venom components with defined biological activities. Indeed, high-throughput Ca^{2+} assays have recently been reported to have been instrumental in the discovery, isolation and characterisation of bioactive peptides from snakes and cone snails.[101–103]

In addition to technical limitations inherent to HTS assays,[100,103] the most notable disadvantage of this strategy for discovery of novel bioactive venom components lies in the high specificity of these assays for their intended molecular target. A lack of activity in such assays does not necessarily imply lack of biological activity, but rather incomplete knowledge of the evolutionary role and thus putative molecular target of such compounds. Accordingly, a multidisciplinary approach, incorporating transcriptomic, evolutionary, and pharmacological analysis will likely be needed to guide the discovery and characterisation of novel venom components with biological activity.

1.3 Case Studies

The multidisciplinary approach described above can yield a wealth of information regarding the evolution of various venom encoding genes. Here, we illustrate this fact with a few examples (Figure 1.4).

1.3.1 Differential Evolution of *Psammophis mossambicus* SVMP Domains[101]

Most viper venoms are haemotoxic in nature, disrupting homeostasis through destruction of red blood cells, coagulation of blood, necrosis, oedema, haemorrhage, *etc.* This results in severe localised and systemic tissue damage and/or organ failure. A subfamily of zinc-dependent matrix-metalloproteinases called the snake venom metalloproteinases (SVMPs) is largely responsible for these pathological effects of viperid snake envenomation. Since the present-day antivenom fails to neutralise these rapidly developing symptoms effectively, a focus of venom research in the recent past has been on studying the effects of SVMP-induced inflammatory reactions.

Snake venom metalloproteinases belong to the reprolysin subfamily of the M12 family of proteinases. They are grouped into three major types based on their domain organisation:[104] (i) P-I SVMPs have only the metalloproteinase domain; (ii) P-II SVMPs contain a disintegrin-like domain in addition to the former; and (iii) P-III SVMPs have an additional cysteine-rich domain. The P-III class of SVMPs have been demonstrated to be more potent in causing haemorrhage than the P-I and P-II classes.[104,105] Hence it is hypothesised that the additional cysteine-rich domain absent from both P-I and P-II SVMPs but present in the P-III SVMP is responsible for the increased potency of the latter. SVMPs have been demonstrated to destroy blood vessels by the degradation of basement membrane proteins and to impair coagulation through fibrinogenolysis. However, several SVMPs devoid of haemorrhagic activity have also been isolated.

The multidomain, multiproduct SVMP gene exhibits some fascinating characteristics such as the selective expression of domains. The selective expression of the disintegrin-like domain from P-II SVMP is well known in viper venoms. Recently, a similar phenomenon was also described in

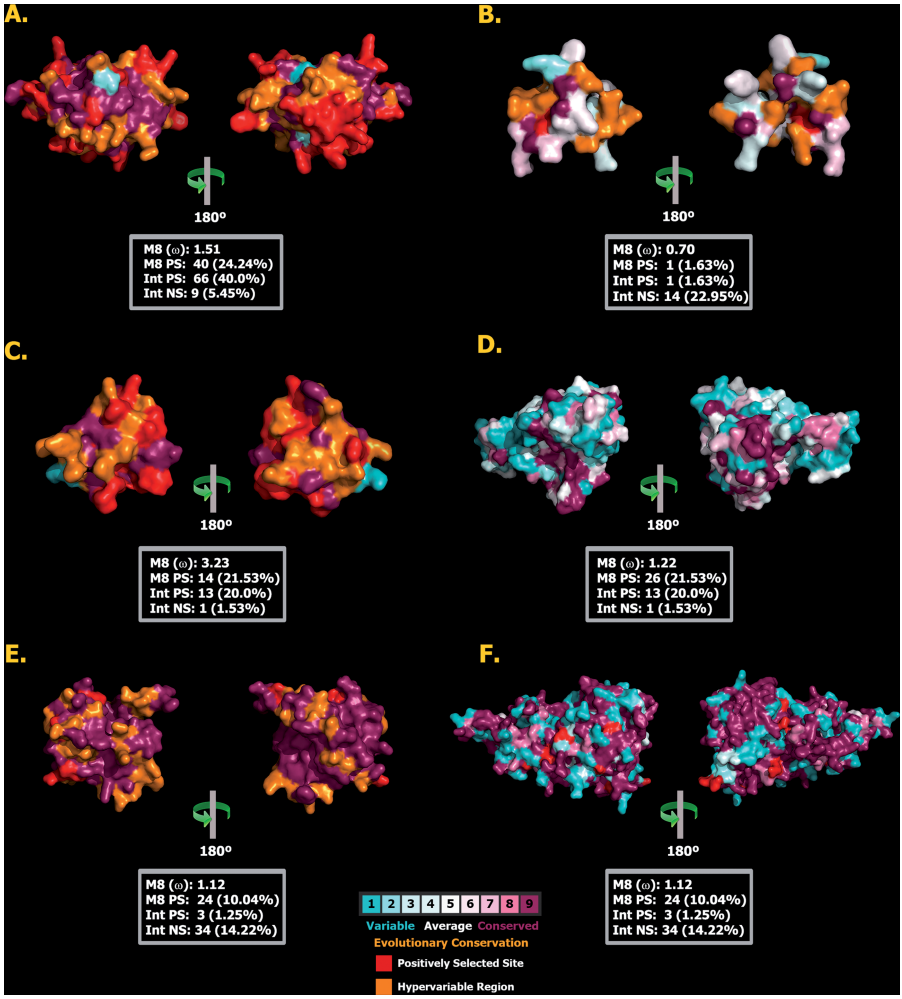


Figure 1.4 Molecular evolution of diverse venom components. (A) Basal snake lectins; (B) Iguania crotamines; (C) Crotalinae crotamines; (D) Elapidae CRiSPs; (E) Homelopsidae veficolins and (F) Vampire bat (*Desmodus rotundus*) plasminogen activator. The homology models show the locations of positively selected sites in red (Model 8, PP $\geq$ 0.95, Bayes-Empirical Bayes approach), along with the site-model 8 computed ω value and the number of positively and negatively selected sites detected using the HyPhy integrative approach [Single Likelihood Ancestor Counting (SLAC), Fixed-Effects Likelihood (FEL), Mixed Effects Model Evolution (MEME): 0.05 significance; Random Effects Likelihood (REL): 50 Bayes factor; Fast, Unconstrained Bayesian AppRoximation (FUBAR): 0.9 posterior probability].

Psammophis mossambicus, in which the P-III SVMP propeptide was shown to be selectively expressed.[41] We have shown that this has put the pre-protein domain under significant selection pressure in these lamprophiid snakes. As a result, the prepro domain of *Psammophis* SVMP has accumulated a pool of mutations at a rapid frequency. Positive selection pressure was found to have influenced evolution of the *Psammophis* monodomain SVMP prepro region more than that of the highly lethal *Echis coloratus* multidomain SVMPs. However, a few species of *Echis* that express similar pre-pro only domains also express the regular multidomain SVMPs. Hence, they do not require similar variations like the *Psammophis*. As a result, they do not accumulate similar variations in the pre-pro region, which remains non-functional. Evolutionary selection analyses conducted on different domains of *Echis* multidomain SVMPs also indicated that the peptidase domain, hypothesised to play a major role in haemotoxicity, accumulates more variations than the disintegrin and cysteine-rich domains. Bioassays revealed that some of the *Psammophis* pre-pro SVMPs inhibit the post-synaptic α7 nicotinic acetylcholine receptor in a similar manner to the α-neurotoxins of Elapidae. Thus, our multidisciplinary approach involving transcriptomics, proteomics, bioassays and bioinformatics allowed us to identify various interesting evolutionary aspects of the *Psammophis* venom arsenal.

1.3.2 Evolution of Vampire Bat Venom[106]

With a level of infamy that extends far beyond the boundaries of science, blood-sucking vampire bats (Chiroptera, Desmodotinae) have been the subject of folk tales, superstitions and stories associated with the legendary Count Dracula for centuries.[107,108] All three species of vampire bat are confined to Central and South America and typically live in caves, tree hollows, and abandoned mines.[109] The relatively rare hairy-legged vampire bat (*Diphylla ecaudata*) feeds exclusively on avian hosts, while the white-winged vampire bat (*Diaemus youngi*) thrives on both mammalian and avian blood, but most likely favours the latter.[107,110] By contrast, the common vampire bat (*Desmodus rotundus*) feeds exclusively on mammals and has established itself in large colonies over an extensive, and apparently widening, distribution.[107,111,112] The expanding population of these bats is attributed to the increasing human population and the associated large number of domesticated animals and livestock, which provide a constant, high-density food supply.[109,112]

All three species of vampire bat are highly specialised for a haematophagous lifestyle. In order to facilitate blood feeding, the vampire bats must be capable of interfering with their prey's natural haemostatic response during both feeding and digestion.[113,114] A typical haemostatic response produces a fibrin clot within minutes of the infliction of a wound, preventing further blood loss. The response commences with the constriction of blood vessels, restricting blood flow to the wound, and is followed by the adhesion of activated platelets to the site of injury and the conversion of fibrinogen to

insoluble fibrin, forming a blood clot.[114] In contrast to this normal response to injury, bleeding from a wound induced by vampire bats may be prolonged from minutes to hours, ensuring a constant flow of blood for the bat to feed upon.[114] The submaxillary gland of vampire bats has been shown to secrete venom that has strong anticoagulant and proteolytic activities.[115] The venom delays the onset of blood clotting by interfering with fibrin formation or acting upon fibrin as it is converted from fibrinogen. In addition it has a strong proteolytic action that breaks up any blood clots that may form. This proteolytic action is accomplished through activation of the host's fibrinolytic system, which converts plasminogen to plasmin, solubilising and removing fibrin clots to prevent excessive fibrin build-up at the site of the wound.

We recently elucidated the evolution of some of the common vampire bat venom encoding genes.[106] By utilising a combined proteomic, transcriptomic, bioinformatics, and phylogenetic approach, we not only discovered a much richer suite of secreted proteins than had been previously recognised, but also unravelled interesting aspects of venom evolution in these majestic flying mammals. For example, it was previously unclear whether the anterior and posterior lobes of the *Desmodus rotundus* submaxillary glands were evolving on different evolutionary trajectories, or if they remained under shared genetic control. Recovery of identical transcripts from the cDNA libraries in this study provides the first evidence that they remain a single expression system. Multiple transcripts of the majority of each protein type were recovered from the cDNA libraries, a pattern consistent with accelerated diversification in toxin multigene families as observed in other venoms.[19,116,117]

1.3.3 Evolution of the Venom Apparatus and Peptide Toxin Characterisation in Terebrid and Turrid Marine Snails

While the progress made in characterising cone snail venom is significant, it is only the tip of the iceberg when considering the extensive biodiversity of venomous marine molluscs. The 700 known species of cone snail are a relatively small component of the biodiversity of venomous molluscs in the superfamily Conoidea, which includes sister groups of the family Terebridae, ~400 species, and Turridae, >10 000 species.[118,119] However, unlike cone snails and their venom, the Terebridae and Turridae have not been as extensively characterised. Recent investigations to characterise terebrid[120–122,147] and turrid[123–125] venom as well as their phylogeny[126–130] suggest that the peptide toxins found in terebrids and turrids, termed teretoxins and turritoxins, respectively, are similar to conopeptides in molecular structure. Namely, teretoxins and turritoxins are expressed as a gene product comprising a signal sequence and an intervening pro-region followed by the mature disulfide-rich peptide toxin. Despite their gene organisation similarity there are several notable differences in size and PTMs between conopeptides, teretoxins, and turritoxins. Teretoxins and turritoxins are

generally larger in size compared with conopeptides. Conopeptides are typically 10–40 amino acid residues long, whereas terebrids and turrids can range up to 70–80 amino acid residues in size. In this respect, teretoxins and turritoxins are similar to peptides toxins from snakes and scorpions. Analysis of terebrid and turrid venom and cDNA transcriptome data further enhance the diversity from conopeptides (Table 1.2). With regard to PTMs, unlike conopeptides and turritoxins, teretoxins do not appear to be extensively post-translationally modified apart from the formation of disulfide bonds. The recent biochemical characterisation of turritoxin cce9a from the turrid species *Crassispira cerithina* revealed a distinctive age-dependent behavioural phenotype when it was injected intracranially into mice.[125] Similar to conantokin G from *Conus geographus*, turritoxin cce9a caused lethargy in mice 12–14 days old and hyperactivity in mice older than 16 days, suggesting that cce9a may target the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptors. However, bioassays of cce9a on NMDA receptors were negative, indicating that while the phenotypic behaviour induced by cce9a is similar to that of conantokin G, the molecular mechanism of action is different. This result is indicative of what has been observed generally in the characterisation of terebrid and turrid venom. While cysteine patterns similar to conopeptides are present in teretoxins and turritoxins, it is not a general indicator for comparative function. Even the disulfide connectivity can vary between homologous toxins from these different Conoidea families. For example, Tv1 from the venom of *Terebra variegata* has an M-like conopeptide arrangement of CC–C–C–CC, but it has a disulfide connectivity unlike any M conopeptide previously described.[147]

Although there are striking similarities to conopeptides, teretoxins and turritoxins have a divergent story to tell in terms of venom composition and functional targets in the nervous system, as is evidenced by the fact that unlike cone snails, not all species of terebrids and turrids hunt prey using a venom apparatus.[120,130] Molecular analyses based on cytochrome c oxidase subunit I (COI) and ribosomal 12S, 16S and 28S genes, combined with anatomical morphometric data, revealed a correlation between the appearance of a venom apparatus and clade delineation in the Terebridae. The Terebridae appear to have lost the venom gland at least eight times during their evolution. The foregut anatomy in the family Terebridae is as varied as the range of variability within the entire superfamily Conoidea. Assuming the diversity of foregut structures in the Terebridae is linked to the diversity of feeding types and prey, it follows that the species diversity of the Terebridae could be linked to the prey diversity, suggesting that terebrids have adapted to dietary changes that may have resulted from predator–prey relationships. As terebrids are largely worm hunters, while cone snails hunt fish, worms, and other molluscs, variations in predator–prey relationships may account for the differences in conopeptides and teretoxins.

The complexity of terebrid and turrid venom is a compelling tale to enhance what has been discovered from cone snails. The possibilities for discovering novel bioactive compounds with unique molecular targets from

Table 1.2 Examples of peptide toxins identified in Turridae and Terebridae venoms

Genus	*Species*	*Peptide toxin*	*Peptide sequence*	*References*
			Turridae	
Crassispira	*cerithina*	Cce9a	GSCGLPCHENRRCGWACYCDDGICKPLRV	125
Gemmula	*diomedea*	Gdm9.1	GDOORFCVHKICYγDSγCNQWCTGGCNOTQGKCDTA	148
Gemmula	*kieneri*	Gkn9.1	GVOORFCRDKNCNγDSγCNOWCTGGCSSVKGNCγS	148
Gemmula	*sogodensis*	Gsg9.1	LDOOGYCTHKICYγDGγCNQWCTAGCNODTGKCDTT	148
Gemmula	*speciosa*	Gsp9a	IDOORYCNHIICYγDSγCSQWCTAGCNSITSKCDT	148
		Gsp9b	GDOORFCRDKLCSGDGDCSVWCTAGCNHDMGKCDTL	148
Lophiotoma	*olangoensis*	OL105	FNTCPGQYDECGNGPEEGECCGTHNYCCKNACTTYHHCHGGRDAGKLLRSLKKLTTH	124
		OL108	QHDCTCVEEGQPCQGQACAAECEYNQCLFDGR	124
		OL11	CLSVCSAEYEPVCGSDGKTYANKCHLMTEACWSPTSITLVHEGKC	124
		OL127	YYICESCWTCESCAGSTESSCVSACNACDLCPNK	124
		OL135	VECNETCEEFCTYCDDNNAEETENCRTDQTDHSRCVDFYTANNLPT	124
		OL139	QEGNVCHRPCFRCHVCGETIAACAACSICIGCEEVVEDACAGNPCYWCDNCG VNDGSHRTTRDTADKTHGGQRDRFFQSIA	124
		OL142	QYCPSNTCNERDECPGNCNDCEYVPGTGPNKRCVKK	124
		OL172	QEECDDPDGLMCCTISEMPTC	124
		OL184	AESCDPYQACVLLSAEGRRVPLCSCAGRDCPNTDSHKIQSMYFCEDVSVVY	124
		OL22	TSCETHQICGRKIIYRDGTTNTQ/EEIDYCRCSGDTDCPKDDVNEISFVD/ EWSYWEVSYYTCL	124
		OL25	SLVCDLECSAEVTTCCETGTCHGITTYNCVGGTEPET	124
		OL38	SEVLECYFECGNWEGTCCDTGICVGIHNCKIPEN	124
		OL47	TSCNAATGRSPGCFCNNDNNCRDTCCPRSDTEKKCTGGPDPCPPRQWPD	124
		OL49	LLCISCCVSITECCQLMSGCAVEIKS	124
		OL55	QACSETSDCLEGLECSGNQCLIPYDGDDSCVIGVGCVYDNGNPGCIRDHRC KGDKKDICTNPATECDEDKVCGYKEGETCYGPCRKGLTCRNTRCQK	124
		OL57	LTCLTKMVECIQLPLDVEDSSDTLCC	124

		OL67	SNTCPGYFGECGDGPEEGECCGMYNYCCKGRCLMLASCQKRRDAGRLL RSLKKLKLTTH	124
		OL71	ASCAARACALSGSNITVTYCSCTDGTTHVCPDGSHETNDMYFCENISGVA	124
		OL78	TQAGNYHYCNEGAAVQPGAGPIIYRQNNVEHPYKMNCRCNPWINKARCCTPGT	124
			Terebridae	
Terebra	*subulata*	Agx-s11a	DCEQHTDCSAASGGPVYCCQDSDCCGGVDYICTNYGQCVRHF	121
		Agx-s6a	SLDEELKSNDCPEYCPHGNECCEHHECRYDPWSRELKCLDS	121
		Agx-s7a	ATNRHQCDTNDDCEEDECCVLVGGNVNNPGVQTRICLACS	121
Hastula	*hectica*	hhe1a	GECCTDCAQTAAANYC	122
		Hhe53	GLSQSGCQAFTGRWCVGCERLRSRVVWECSPKRVVNSI	122
		Hhe6.1	GMGIGINLPPCIKNGEYCNPWTGSIILGGACCGTCTDYECH	122
		Hhe6.2	ALPCPYGCPLRCCHMTDGVCLRNKQGC	122
		Hhe6.3	VLFTPPELLGCGNRCSDDCCKWGRCQPGCTD	122
		Hhe6.4	SSLHCGDDPWCPTGCCENEDCDIGCKRDWEKSRSQP	122
		hhe7a	ARCEQCPSYCCQSDSPPECDGCE	122
		Hhe9	ALSVLLQUSCTMCLFCCYL	122
		Hhe9.2	DEEVGCFPNVCKNDGNCSIETSTGMTRCQCLEGYTGHVCENPL	122
		hhe9a	YEENCGTEYCTSKIGCPGRCVCKEYNYNGEITRRCRA	122
		hheTx1	DCLPCGHDVCC	122
		hheTx2	SCSSGCSDCNSDDSCQCTLNQFTNSDSCCC	122
		hheTx3	KQCTSNMCSADCSPGCCIIDKLEWCTCDC	122
		hheTx4	NEVCPPGECQQYCCDLRKCKCINLSFYGLTCNCDS	122
		hheTx5	NEVCPPGRCEPYCCDPRKCKCLSIDFYGLVCNCDS	122
Terebra	*variegata*	Tv1	TR(I/L)CCGCYWNGSKDVCSQSCC	147

terebrid and turrid snails is a promising scientific exploration that requires an integrated approach. A strategy combining genomic, proteomic, and transcriptomic data is not only a promising way to identify the factors that led to diversification of the Terebridae and Turridae, but it also advances the characterisation of terebrid and turrid peptide toxins with novel function and potentially new therapeutic applications.

1.4 Significance of the Combined Approach

1.4.1 Mutation of the Surface Chemistry

Bioinformatics and evolutionary analyses have shown that most of the mutations that accumulate within venom components affect changes on the molecular surface (Figure 1.5).[131–133] Mutation of the surface chemistry ensures that the enzymatic activity, when present, is conserved while variation is still generated in residues responsible of interaction with different target cells and receptors in the prey. This results in the generation of exquisite variation,

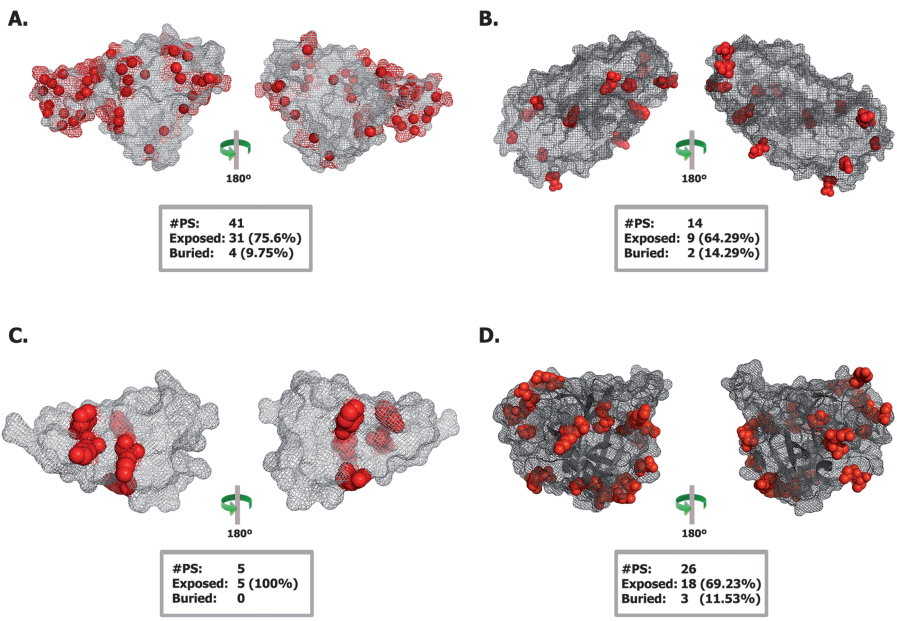


Figure 1.5 Surface accessibility of venom components. A depiction of locations of positively selected sites in different toxins and their surface accessibility: (A) Caenophidian snake CRiSPs; (B) *Crotalus oreganus helleri* L-amino-acid oxidase; (C) Vampire Bat (*Desmodus rotundus*) kunitz protease inhibitor domain I and (D) Coleoid serine protease. Total number of positively selected sites (#PS) detected by PAML (Model 8, Bayes Empirical Bayes approach, Posterior probability $\geq$ 0.95), total number of exposed and buried positively selected sites are also indicated.

utilising the existing venom arsenal. Identification of such phenomena associated with toxin evolution is extremely important in drug design and antivenom production. Antivenom can be targeted at the conserved structural residues, thus efficiently neutralising the severe effects of envenomation.

1.4.2 Alternative (Differential) Splicing

Alternative splicing is a post-transcriptional process that involves the modification and rearrangement of exons, resulting in multiple protein products (isoforms) encoded by the same genetic sequence. It is the most common phenomenon through which eukaryotes generate proteomic complexity and diversity. Alternative splicing can reduce the maintenance cost associated with multiple-gene families and at the same time increase the genetic potential for encoding complex proteomes. For instance, the *Drosophila melanogaster* DSCAM gene alone can code for a hypothetical maximum of approximately 38 106 peptides utilising the same stretch of nucleotides.[134] Molecular evolution research has shown that several venom-encoding genes adopt alternative splicing to generate an array of toxin types, without increasing the number of copies of the gene itself.[135,136] These isoforms can target a wide range of cells and receptors[137] and could potentially acquire completely different biochemical functions from one another.

1.4.3 Post-Translational Modification

To further increase the variation of the venom arsenal, many venom components undergo PTMs such as glycosylation, phosphorylation, disulfide bridge formation, proteolysis, *etc.* PTMs are also important for achieving a functional state, stabilisation, proper folding of the mature protein, protection from lytic enzymes of the target animal that may disrupt the venom component's function, and exposure of hidden residues that could enable the recognition of new biological targets. PTMs are essential for the modulation of biochemical activities. For instance, glycosylation of the *Naja kouthia* 3FTx decreases its cytotoxicity and expression by two-fold, relative to the unglycosylated form.[138] PTMs can only be detected through proteomic analyses.

1.5 Concluding Remarks

The last two decades has seen a surge in projects exploiting the extraordinary biological potency of venom components to develop novel drugs and diagnostics for human diseases, or as probes to study cells and their receptors.[3,139] Encouraged by the substantial medicinal and fiscal success of the Bristol-Myers Squibb angiotensin-converting enzyme inhibitor, captopril,[140] many pharmaceutical companies have invested in venom-based drug discovery programs.[3] The majority of the currently approved products were developed from snake venom proteins with distinct cardiovascular specificities, particularly those that target thrombin, fibrinogen, and integrin

receptors.[3,141] Rapid advances in proteomics, genomics, and transcriptomics have since resulted in affordable technology platforms[100,142–144] that enable mining of venom proteins/peptides for drug discovery from species which, unlike snakes, produce venom in very small quantities. For example, the toxin repertoires of spiders are estimated to contain more than 10 million compounds available for bioprospecting.[142] These reports illustrate that venom peptides, particularly those enriched with the molecular stability imparted by extensive disulfide bonds, are driving the development of, amongst other things, new analgesics, anti-tumour agents and even insecticides. Drug bioprospecting activity is likely to continue to rise as largely unstudied venomous animal lineages are exploited for discovery of novel lead compounds. Venoms are now beginning to receive a great deal of attention as natural sources of novel diagnostic and therapeutic compounds.[140] The venom pool studied to date, often with particular focus on certain toxin types through selective assaying, represents an infinitesimally small representation of the true diversity available. An understanding of the evolutionary and ecological biology of different venomous animal lineages is therefore fundamental to the optimal selection of biological targets for future drug discovery programs.

Acknowledgements

BGF was funded by an Australian Research Council Future Fellowship and by The University of Queensland. This research was supported in part by Fundação para a Ciência e a Tecnologia (FCT) through the PhD grant conferred to KS (SFRH/BD/61959/2009). EABU would like to acknowledge funding from the University of Queensland (International Postgraduate Research Scholarship, UQ Centennial Scholarship, and UQ Advantage Top-Up Scholarship) and the Norwegian State Education Loans Fund. SAA was the recipient of a postdoctoral fellowship (PDRF Phase II Batch-V) from the Higher Education Commission (HEC Islamabad) of Pakistan. NRC was funded by the UK Natural Environment Research Council. MH acknowledges funding from U.S. National Science Foundation (NSF-CHE 1247550) and Alfred P. Sloan Foundation (B2010-37).

References

1. D. J. Newman, G. M. Cragg, S. Holbeck and E. A. Sausville, *Curr. Cancer Drug Targets,* 2002, **2**, 279–308.
2. R. J. Lewis and M. L. Garcia, *Nat. Rev. Drug Discovery,* 2003, **2**, 790–802.
3. G. F. King, *Expert Opin. Biol. Ther.,* 2011, **11**, 1469–1484.
4. B. G. Fry, K. Roelants, D. E. Champagne, H. Scheib, J. D. A. Tyndall, G. F. King, T. J. Nevalainen, J. A. Norman, R. J. Lewis, R. S. Norton, C. Renjifo and R. C. Rodríguez de la Vega, *Annu. Rev. Genomics Hum. Genet.,* 2009, **10**, 483–511.
5. E. Gong, L. D. Martin, D. A. Burnham and A. R. Falk, *Proc. Natl. Acad. Sci. U.S.A.,* 2010, **107**, 766–768.

6. J. E. Church and W. C. Hodgson, *Toxicon,* 2002, **40**, 1083–1093.
7. D. C. de Graaf, M. Aerts, E. Danneels and B. Devreese, *J. Proteomics,* 2009, **72**, 145–154.
8. N. Peiren, F. Vanrobaeys, D. C. de Graaf, B. Devreese, J. Van Beeumen and F. J. Jacobs, *Biochim. Biophys. Acta,* 2005, **1752**, 1–5.
9. L. Van Valen, *Evol. Theor.,* 1973, **1**, 1–30.
10. S. P. Mackessy, *Toxicon,* 2010, **55**, 1463–1474.
11. N. Jorge da Silva, Jr. and S. D. Aird, *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.,* 2001, **128**, 425–456.
12. A. Barlow, C. E. Pook, R. A. Harrison and W. Wüster, *Proc. Biol. Sci.,* 2009, **276**, 2443–2449.
13. H. L. Gibbs and S. P. Mackessy, *Toxicon,* 2009, **53**, 672–679.
14. J. L. Glenn and R. Straight, *Toxicon,* 1978, **16**, 81–84.
15. J. J. Calvete, L. Sanz, P. Cid, P. de la Torre, M. Flores-Diaz, M. C. Dos Santos, A. Borges, A. Bremo, Y. Angulo, B. Lomonte, A. Alape-Giron and J. M. Gutierrez, *J. Proteome Res.,* 2010, **9**, 528–544.
16. M. C. Menezes, M. F. Furtado, S. R. Travaglia-Cardoso, A. C. Camargo and S. M. Serrano, *Toxicon,* 2006, **47**, 304–312.
17. D. V. Andrade and A. S. Abe, *Herpetologica,* 1999, **55**, 200–204.
18. S. K. Sutherland and J. Tibballs, *Australian animal toxins: the creatures, their toxins and care of the poisoned patient*, Oxford University Press, 2001.
19. B. G. Fry, *Genome Res.,* 2005, **15**, 403–420.
20. T. C. Assumpcao, J. M. Ribeiro and I. M. Francischetti, *Toxins,* 2012, **4**, 296–322.
21. B. G. Fry, W. Wüster, R. M. Kini, V. Brusic, A. Khan, D. Venkataraman and A. P. Rooney, *J. Mol. Evol.,* 2003, **57**, 110–129.
22. J. A. Veenstra, *Arch. Insect Biochem. Physiol.,* 2000, **43**, 49–63.
23. B. G. Fry, K. Winter, J. A. Norman, K. Roelants, R. J. Nabuurs, M. J. van Osch, W. M. Teeuwisse, L. van der Weerd, J. E. McNaughtan, H. F. Kwok, H. Scheib, L. Greisman, E. Kochva, L. J. Miller, F. Gao, J. Karas, D. Scanlon, F. Lin, S. Kuruppu, C. Shaw, L. Wong and W. C. Hodgson, *Mol. Cell. Proteomics,* 2010, **9**, 2369–2390.
24. M. Herrera, J. Fernandez, M. Vargas, M. Villalta, A. Segura, G. Leon, Y. Angulo, O. Paiva, T. Matainaho, S. D. Jensen, K. D. Winkel, J. J. Calvete, D. J. Williams and J. M. Gutierrez, *J. Proteomics,* 2012, **75**, 2128–2140.
25. B. G. Fry, H. Scheib, L. van der Weerd, B. Young, J. McNaughtan, S. F. Ramjan, N. Vidal, R. E. Poelmann and J. A. Norman, *Mol. Cell. Proteomics,* 2008, **7**, 215–246.
26. B. G. Fry, N. G. Lumsden, W. Wüster, J. C. Wickramaratna, W. C. Hodgson and R. M. Kini, *J. Mol. Evol.,* 2003, **57**, 446–452.
27. J. Pawlak, S. P. Mackessy, N. M. Sixberry, E. A. Stura, M. H. Le Du, R. Ménez, C. S. Foo, A. Ménez, S. Nirthanan and R. M. Kini, *FASEB J.,* 2009, **23**, 534–545.
28. J. Pawlak, S. P. Mackessy, B. G. Fry, M. Bhatia, G. Mourier, C. Fruchart-Gaillard, D. Servent, R. Ménez, E. Stura, A. Ménez and R. M. Kini, *J. Biol. Chem.,* 2006, **281**, 29030–29041.

29. M. Goyffon and C. Landon, *C. R. Seances Soc. Biol. Ses Fil.,* 1998, **192**, 445–462.
30. H. Rochat, P. Bernard and F. Couraud, *Adv. Cytopharmacol.,* 1979, **3**, 325–334.
31. V. Herzig and G. F. King, in *Spider Ecophysiology*, ed. W. Nentwig, Springer Berlin Heidelberg, 2013, pp. 203–215.
32. J. A. Coddington and H. W. Levi, *Annu. Rev. Ecol. Systemat.,* 1991, **22**, 565–592.
33. L. I. Anderson and N. H. Trewin, *Palaeontology,* 2003, **46**, 467–509.
34. N. R. Casewell, W. Wüster, F. J. Vonk, R. A. Harrison and B. G. Fry, *Trends Ecol. Evol.*, 2013, **28**, 219–229.
35. E. A. Undheim, D. N. Georgieva, H. H. Thoen, J. A. Norman, J. Mork, C. Betzel and B. G. Fry, *Toxicon,* 2010, **56**, 897–913.
36. I. Koludarov, K. Sunagar, E. A. Undheim, T. N. Jackson, T. Ruder, D. Whitehead, A. C. Saucedo, G. R. Mora, A. C. Alagon, G. F. King, A. Antunes and B. G. Fry, *J. Mol. Evol.,* 2012, **75**, 168–183.
37. B. G. Fry, K. Roelants and J. A. Norman, *J. Mol. Evol.,* 2009, **68**, 311–321.
38. N. Vidal and S. B. Hedges, *C. R. Biol.,* 2005, **328**, 1000–1008.
39. B. G. Fry, N. Vidal, J. A. Norman, F. J. Vonk, H. Scheib, S. F. R. Ramjan, S. Kuruppu, K. Fung, S. B. Hedges, M. K. Richardson, W. C. Hodgson, V. Ignjatovic, R. Summerhayes and E. Kochva, *Nature,* 2006, **439**, 584–588.
40. B. G. Fry, K. Roelants, K. Winter, W. C. Hodgson, L. Griesman, H. F. Kwok, D. Scanlon, J. Karas, C. Shaw, L. Wong and J. A. Norman, *Mol. Biol. Evol.,* 2010, **27**, 395–407.
41. B. G. Fry, H. Scheib, L. van der Weerd, B. Young, J. McNaughtan, S. F. Ryan Ramjan, N. Vidal, R. E. Poelmann and J. A. Norman, *Mol. Cell. Proteomics,* 2008, **7**, 215–246.
42. L. Junqueira-de-Azevedo Ide and P. L. Ho, *Gene,* 2002, **299**, 279–291.
43. S. C. Wagstaff and R. A. Harrison, *Gene,* 2006, **377**, 21–32.
44. R. S. Rodrigues, J. Boldrini-Franca, F. P. Fonseca, P. de la Torre, F. Henrique-Silva, L. Sanz, J. J. Calvete and V. M. Rodrigues, *J. Proteomics,* 2012, **75**, 2707–2720.
45. D. R. Rokyta, A. R. Lemmon, M. J. Margres and K. Aronow, *BMC Genomics,* 2012, **13**, 312.
46. Y. Terrat, D. Biass, S. Dutertre, P. Favreau, M. Remm, R. Stocklin, D. Piquemal and F. Ducancel, *Toxicon,* 2012, **59**, 34–46.
47. S. Gotz, J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talon, J. Dopazo and A. Conesa, *Nucleic Acids Res.,* 2008, **36**, 3420–3435.
48. S. Gotz, R. Arnold, P. Sebastian-Leon, S. Martin-Rodriguez, P. Tischler, M. A. Jehl, J. Dopazo, T. Rattei and A. Conesa, *Bioinformatics,* 2011, **27**, 919–924.
49. A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles, *Bioinformatics,* 2005, **21**, 3674–3676.
50. A. Conesa and S. Gotz, *Int. J. Plant Genomics,* 2008, **2008**, 619832.

51. N. R. Casewell, R. A. Harrison, W. Wüster and S. C. Wagstaff, *BMC Genomics,* 2009, **10**, 564.
52. D. R. Rokyta, K. P. Wray, A. R. Lemmon, E. M. Lemmon and S. B. Caudle, *Toxicon,* 2011, **57**, 657–671.
53. T. Dagan, *Trends Microbiol.,* 2011, **19**, 483–491.
54. S. Halary, J. W. Leigh, B. Cheaib, P. Lopez and E. Bapteste, *Proc. Natl. Acad. Sci. U.S.A.,* 2010, **107**, 127–132.
55. F. Jungo and A. Bairoch, *Toxicon,* 2005, **45**, 293–301.
56. M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang and T. Ideker, *Bioinformatics,* 2011, **27**, 431–432.
57. M. Bastian, S. Heymann and M. Jacomy, *International AAAI Conference on Weblogs and Social Media*, 2009.
58. G. E. Crooks, G. Hon, J. M. Chandonia and S. E. Brenner, *Genome Res.,* 2004, **14**, 1188–1190.
59. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.,* 1997, **25**, 3389–3402.
60. D. G. Higgins and P. M. Sharp, *Gene,* 1988, **73**, 237–244.
61. R. C. Edgar, *Nucleic Acids Res.,* 2004, **32**, 1792–1797.
62. F. Ronquist and J. P. Huelsenbeck, *Bioinformatics,* 2003, **19**, 1572–1574.
63. S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel, *Syst. Biol.,* 2010, **59**, 307–321.
64. A. Stamatakis, P. Hoover and J. Rougemont, *Syst. Biol.,* 2008, **57**, 758–771.
65. D. Posada and K. A. Crandall, *J. Mol. Evol.,* 2002, **54**, 396–402.
66. G. A. McVean, *Heredity (Edinb.),* 2001, **87**, 613–620.
67. M. Worobey, *Mol. Biol. Evol.,* 2001, **18**, 1425–1434.
68. K. S. Lole, R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard and S. C. Ray, *J. Virol.,* 1999, **73**, 152–160.
69. D. Martin and E. Rybicki, *Bioinformatics,* 2000, **16**, 562–563.
70. I. Milne, F. Wright, G. Rowe, D. F. Marshall, D. Husmeier and G. McGuire, *Bioinformatics,* 2004, **20**, 1806–1807.
71. S. D. Frost, Y. Liu, S. L. Pond, C. Chappey, T. Wrin, C. J. Petropoulos, S. J. Little and D. D. Richman, *J. Virol.,* 2005, **79**, 6523–6527.
72. W. Delport, A. F. Poon, S. D. Frost and S. L. Kosakovsky Pond, *Bioinformatics,* 2010, **26**, 2455–2457.
73. S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk and S. D. Frost, *Mol. Biol. Evol.,* 2006, **23**, 1891–1901.
74. M. Kimura, Cambridge University Press, 1983, Cambridge.
75. N. Goldman and Z. Yang, *Mol. Biol. Evol.,* 1994, **11**, 725–736.
76. Z. Yang, *Mol. Biol. Evol.,* 1998, **15**, 568–573.
77. Z. Yang, *Mol. Biol. Evol.,* 2007, **24**, 1586–1591.
78. S. L. Pond and S. D. Frost, *Mol. Biol. Evol.,* 2005, **22**, 478–485.
79. S. L. Kosakovsky Pond, B. Murrell, M. Fourment, S. D. Frost, W. Delport and K. Scheffler, *Mol. Biol. Evol.,* 2011, **28**, 3033–3043.

80. K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics,* 2006, **22**, 195–201.
81. F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli and T. Schwede, *Nucleic Acids Res.,* 2009, **37**, D387–D392.
82. M. C. Peitsch, *Biochem. Soc. Trans.,* 1996, **24**, 274–279.
83. L. A. Kelley and M. J. Sternberg, *Nat. Protoc.,* 2009, **4**, 363–371.
84. A. Armon, D. Graur and N. Ben-Tal, *J. Mol. Biol.,* 2001, **307**, 447–463.
85. A. Doron-Faigenboim, A. Stern, I. Mayrose, E. Bacharach and T. Pupko, *Bioinformatics,* 2005, **21**, 2101–2103.
86. A. Stern, A. Doron-Faigenboim, E. Erez, E. Martz, E. Bacharach and T. Pupko, *Nucleic Acids Res.,* 2007, **35**, W506–W511.
87. J. Davis, A. Jones and R. J. Lewis, *Peptides,* 2009, **30**, 1222–1227.
88. P. Escoubas, B. L. Sollod and G. F. King, *Toxicon,* 2006, **47**, 650–663.
89. B. M. Olivera and L. J. Cruz, *Toxicon,* 2001, **39**, 7–14.
90. B. M. Olivera, J. M. McIntosh, L. J. Cruz, F. A. Luque and W. R. Gray, *Biochemistry,* 1984, **23**, 5087–5090.
91. C. Clark, B. M. Olivera and L. J. Cruz, *Toxicon,* 1981, **19**, 691–699.
92. E. W. McCleskey, A. P. Fox, D. H. Feldman, L. J. Cruz, B. M. Olivera, R. W. Tsien and D. Yoshikami, *Proc. Natl. Acad. Sci. U.S.A.,* 1987, **84**, 4327–4331.
93. G. P. Miljanich, *Curr. Med. Chem.,* 2004, **11**, 3029–3040.
94. R. J. Lewis, S. Dutertre, I. Vetter and M. J. Christie, *Pharmacol. Rev.,* 2012, **64**, 259–298.
95. L. J. Greene, A. C. Camargo, E. M. Krieger, J. M. Stewart and S. H. Ferreira, *Circ. Res.,* 1972, **31**(Suppl. 2), 62–71.
96. S. H. Ferreira, *Br. J. Pharmacol. Chemother.,* 1965, **24**, 163–169.
97. M. A. Ondetti, B. Rubin and D. W. Cushman, *Science,* 1977, **196**, 441–444.
98. I. A. Sharpe, J. Gehrmann, M. L. Loughnan, L. Thomas, D. A. Adams, A. Atkins, E. Palant, D. J. Craik, D. J. Adams, P. F. Alewood and R. J. Lewis, *Nat. Neurosci.,* 2001, **4**, 902–907.
99. J. H. Zhang, T. D. Chung and K. R. Oldenburg, *J. Biomol. Screen.,* 1999, **4**, 67–73.
100. I. Vetter, J. L. Davis, L. D. Rash, R. Anangi, M. Mobli, P. F. Alewood, R. J. Lewis and G. F. King, *Amino Acids,* 2011, **40**, 15–28.
101. A. Brust, K. Sunagar, E. A. Undheim, I. Vetter, D. Yang, N. R. Casewell, T. Ruder, T. N. Jackson, I. Koludarov, P. F. Alewood, W. C. Hodgson, R. J. Lewis, G. F. King, A. Antunes, I. Hendrikx and B. G. Fry, *Mol. Cell. Proteomics,* 2012, **12**, 651–663.
102. F. Marcon, M. Leblanc, I. Vetter, R. J. Lewis, P. Escoubas and G. M. Nicholson, *Biochem. Pharmacol.,* 2012, **84**, 851–863.
103. I. Vetter, *Adv. Exp. Med. Biol.,* 2012, **740**, 45–82.
104. L. A. Hite, L. G. Jia, J. B. Bjarnason and J. W. Fox, *Arch. Biochem. Biophys.,* 1994, **308**, 182–191.
105. J. B. Bjarnason and J. W. Fox, *Pharmacol. Ther.,* 1994, **62**, 325–372.

106. D. H. W. Low, K. Sunagar, E. A. B. Undheim, S. A. Ali, A. C. Alagon, T. Ruder, T. N. W. Jackson, S. P. Gonzalez, G. F. King, A. Jones, A. Antunes and B. G. Fry, *J. Proteomics*, 2013, **89**, 95–111.
107. M. D. Tuttle in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 1–6.
108. C. B. Villa and R. M. Canela in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 233–240.
109. C. E. Ray, O. J. Linares and G. S. Morgan in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 19–30.
110. K. F. Koopman in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 7–18.
111. K. P. Bhatnagar in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 4–70.
112. A. M. Greenhall in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 111–132.
113. A. V. Basanova, I. P. Baskova and L. L. Zavalova, *Biochemistry (Mosc.),* 2002, **67**, 143–150.
114. C. M. Hawkey in *Natural history of vampire bats*, ed. A. M. Greenhall and U. Schmidt, CRC Press, USA, 1988, pp. 133–142.
115. P. E. Disanto, *J. Morphol.,* 1960, **106**, 301–335.
116. T. F. Duda, Jr. and S. R. Palumbi, *Proc. Natl. Acad. Sci. U.S.A.,* 1999, **96**, 6820–6823.
117. B. G. Fry, W. Wüster, R. M. Kini, V. Brusic, A. Khan, D. Venkataraman and A. P. Rooney, *J. Mol. Evol.,* 2003, **57**, 110–129.
118. P. Bouchet, P. Lozouet, P. Maestrati and V. Heros, *Biol. J. Linnean Soc.,* 2002, **75**, 421–436.
119. P. Bouchet, Y. I. Kantor, A. Sysoev and N. Puillandre, *J. Molluscan Stud.,* 2011, **77**, 273–308.
120. N. Puillandre and M. Holford, *BMC Chem. Biol.,* 2010, **10**, 7.
121. J. S. Imperial, M. Watkins, P. Chen, D. R. Hillyard, L. J. Cruz and B. M. Olivera, *Toxicon,* 2003, **42**, 391–398.
122. J. S. Imperial, Y. Kantor, M. Watkins, F. M. Heralde, 3rd, B. Stevenson, P. Chen, K. Hansson, J. Stenflo, J. P. Ownby, P. Bouchet and B. M. Olivera, *J. Exp. Zool., Part B.,* 2007, **308**, 744–756.
123. E. Lopez-Vera, E. P. Heimer de la Cotera, M. Maillo, J. R. Riesgo-Escovar, B. M. Olivera and M. B. Aguilar, *Toxicon,* 2004, **43**, 365–374.
124. M. Watkins, D. R. Hillyard and B. M. Olivera, *J. Mol. Evol.,* 2006, **62**, 247–256.
125. A. B. Cabang, J. S. Imperial, J. Gajewiak, M. Watkins, P. S. Corneli, B. M. Olivera and G. P. Concepcion, *Toxicon,* 2011, **58**, 672–680.
126. N. Puillandre, S. Samadi, M. C. Boisselier, A. V. Sysoev, Y. I. Kantor, C. Cruaud, A. Couloux and P. Bouchet, *Mol. Phylogenet. Evol.,* 2008, **47**, 1122–1134.
127. M. Holford, N. Puillandre, Y. Terryn, C. Cruaud, B. Olivera and P. Bouchet, *Mol. Biol. Evol.,* 2009, **26**, 15–25.

128. M. Holford, N. Puillandre, M. V. Modica, M. Watkins, R. Collin, E. Bermingham and B. M. Olivera, *PLoS One,* 2009, **4**, e7667.
129. N. Puillandre, C. P. Meyer, P. Bouchet and B. M. Olivera, *Zool. Scr.,* 2011, **40**, 350–363.
130. M. Castelin, N. Puillandre, Y. I. Kantor, M. V. Modica, Y. Terryn, C. Cruaud, P. Bouchet and M. Holford, *Mol. Phylogenet. Evol.,* 2012, **64**, 21–44.
131. K. Sunagar, W. E. Johnson, S. J. O'Brien, V. Vasconcelos and A. Antunes, *Mol. Biol. Evol.,* 2012, **29**, 1807–1822.
132. R. M. Kini and Y. M. Chan, *J. Mol. Evol.,* 1999, **48**, 125–132.
133. N. R. Casewell, S. C. Wagstaff, R. A. Harrison, C. Renjifo and W. Wüster, *Mol. Biol. Evol.,* 2011, **28**, 2637–2649.
134. D. L. Black, *Cell,* 2000, **103**, 367–370.
135. T. J. Fujimi, T. Nakajyo, E. Nishimura, E. Ogura, T. Tsuchiya and T. Tamiya, *Gene,* 2003, **313**, 111–118.
136. T. Tamiya, S. Ohno, E. Nishimura, T. J. Fujimi and T. Tsuchiya, *Toxicon,* 1999, **37**, 181–185.
137. D. Servent, V. Winckler-Dietrich, H. Y. Hu, P. Kessler, P. Drevet, D. Bertrand and A. Ménez, *J. Biol. Chem.,* 1997, **272**, 24279–24286.
138. A. V. Osipov, M. V. Astapova, V. I. Tsetlin and Y. N. Utkin, *Eur. J. Biochem.,* 2004, **271**, 2018–2027.
139. C. J. Bohlen, A. T. Chesler, R. Sharif-Naeini, K. F. Medzihradszky, S. Zhou, D. King, E. E. Sanchez, A. L. Burlingame, A. I. Basbaum and D. Julius, *Nature,* 2011, **479**, 410–414.
140. D. W. Cushman and M. A. Ondetti, *Hypertension,* 1991, **17**, 589–592.
141. C. Y. Koh and R. M. Kini, *Toxicon,* 2012, **59**, 497–506.
142. P. Escoubas and G. F. King, *Expert Rev. Proteomics,* 2009, **6**, 221–224.
143. B. M. Olivera, D. R. Hillyard, M. Marsh and D. Yoshikami, *Trends Biotechnol.,* 1995, **13**, 422–426.
144. S. Liang, *Expert Rev. Proteomics,* 2008, **5**, 731–746.
145. E. Pennisi, *Science,* 2003, **300**, 1694.
146. S. Halary, J. O. McInerney, P. Lopez and E. Bapteste, *BMC Evol. Biol.,* 2013, **13**, 146.
147. P. Anand, A. Grigoryan, M. H. Bhuiyan, B. Ueberheide, V. Russell, J. Quinoñez, P. Moy, B. T. Chait, S. F. Poget and M. Holford, *PloS One,* 2014, **9**, e94122.
148. Heralde III and M. Francisco, *et al.*, *Toxicon,* 2008, **51**, 890–897.