

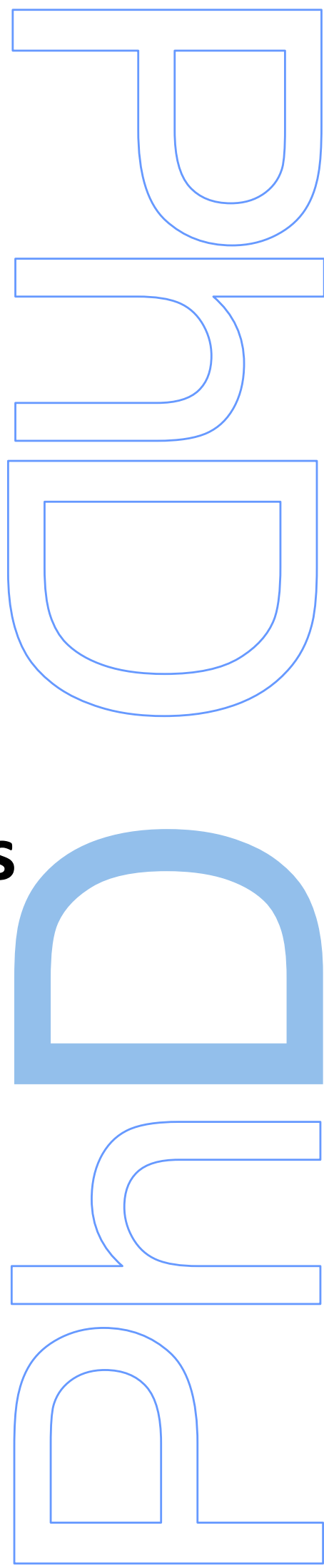
Proteomics and Transcriptomics of Venomous Animals

Dany Domínguez Pérez

PhD Thesis presented to the
Faculty of Sciences of the University of Porto

Biology

2017



Proteomics and Transcriptomics of Venomous Animals

Dany Domínguez Pérez

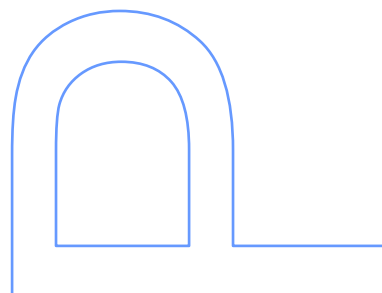
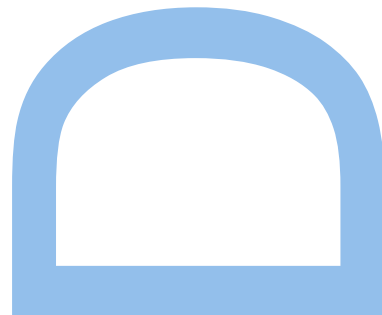
Biology
Faculty of Sciences
2017

Supervisor

Prof. Agostinho Antunes,
Faculty of Sciences of the University of Porto

Co-supervisor

Prof. Vitor Vasconcelos,
Faculty of Sciences of the University of Porto



"Look deep into nature, and then you will understand everything better"

Albert Einstein

Dedicated to my son and my wife

Acknowledgements

I acknowledge the Portuguese Fundação para a Ciência e a Tecnologia (FCT) for financial support of my PhD project (SFRH/BD/80592/2011). This study was funded in part by the Strategic Funding UID/Multi/04423/2013 through national funds provided by FCT and the European Regional Development Fund (ERDF) in the framework of the program PT2020, by the European Structural and Investment Funds (ESIF) through the Competitiveness and Internationalization Operational Program—COMPETE 2020 and by National Funds through the FCT under the project PTDC/AAG-GLO/6887/2014 (POCI-01-0124-FEDER-016845), and by the Structured Programs of R&D&I INNOVMAR—Innovation and Sustainability in the Management and Exploitation of Marine Resources (NORTE-01-0145-FEDER-000035, Research Line NOVELMAR), CORAL NORTE (NORTE-01-0145-FEDER-000036), and MarInfo—Integrated Platform for Marine Data Acquisition and Analysis (NORTE-01-0145-FEDER-000031), and funded by the Northern Regional Operational Program (NORTE2020) through the ERDF.

I want to give a special acknowledgment to my Supervisor Prof. Agostinho Antunes and Co-supervisor Prof. Vítor Vasconcelos for letting me come to work with them and their teams, for providing me all the resources needed in the execution of the work presented in this thesis, for always show kindness, respect, and trust.

Thanks to my colleagues from EGB and BBE, CIIMAR, for facilitating me the integration inside the groups; specially to Guillermin Aguero Chapin for his orientation during the process, to Bárbara Frazão, Tibisay Escalona, Imran Khan, Jorge Neves, Micaela Vale, João Morais, Pratheepa Moorthy, Dina Gomes, Aldo Barreiro, Sofia Costa, Raquel Castelo Branco, Margarida Costa, Vitor Ramos, Pedro Leão, Cristiana Moreira, Marisa Silva, Anoop Alex and Cidália Gomes, for their useful help at the beginning.

To those colleagues who participated directly in the research work with me: Carlos Manlio Diaz Garcia, Neivys García Delgado, Yusvel Sierra Gómez, Olga Castañeda, Carlos Varela, Armando Alexei Rodríguez, Hugo Osorio, Joana Azevedo, Alexandre Campos, Maria V Turkina, Tiago Ribeiro, Isabel Cunha, Ralph Urbatzka, Jordi Durban, Javier Torres Lopez, Reinaldo Molina Ruiz, Tito Mendes, Emanuel Maldonado, Filipe Silva, Bruno Reis and Juan J. Calvete.

Thanks to SASUP for provide me a comfortable lodging, specially to the workers of the residence Residência Universitária Campo Alegre 695, (Pólo III): Angela Braga, and for give me the opportunity to meet people from many countries and culture who shared their experiences, which has allowed me to understand the story much better and leave as a result many bonds of friendship worldwide, only real and unique reason to feel proud.

To those colleagues who helped in the thesis preparation: Daniela Almeida and João Paulo Machado.

Thanks to faculty staff, specially to Rosária Seabra and Ana Isabel Barreira.

I want to thank those special people who were always there, who helped me many times or who welcomed me in their family circle at least once: Fernando Cagide Fajin and José Luis Cagide Fajin Bros, Doris Decoro Rojas and his husband and great Portuguese friend Antonio Luis Lopes de Sousa Castro, Robert Carcasses and Yuselis Castaño, Yonni Romaguera and Lisa Benamati, Rudy and Yaya, Gerardo González jr, Tibysay Escalona, Alexandre Campos, Quiaoquio Chen and Carlos Gustavo Moraes Castro.

Finally, thanks to all my family member and relatives:

To my maternal grandparents who represent "the theory of everything", Reymundo Pérez (I hope you feel happy and a bit proud wherever you are) and Inocencia Rodríguez: the effort, Altruism, gentleness and humility.

To my mother Maritza Pérez Rodríguez for the education and for giving me her infinite strength to overcome all obstacles.

To my sister Mayté Domínguez and my brother-in-law Héctor González, because all the help and support they have given me.

Specially, to my wife Yudermys Moya Chaviano and to my son Eiden Fabián Domínguez Moya, thank you for all the love, understanding, affection, for the endless sacrifice of watching time pass, while we stay away from each other. All this work is dedicated to you. The fruit harvested is yours, and if it would produce more, it will also be yours. I just hope that the knowledge and experiences that come to our home enrich our daily life and trace the path for my son to the truth.

Thanks to Portugal and its people for welcoming me throughout this period which has been a great experience.



Abstract

The presence of toxins is a feature that confers significant advantages to venomous animals in the struggle for survival. Throughout evolution, many group of animals have been independently developing specialized tissues coupled to a delivery system like fangs, needles, harpoons, to produce and inoculate venoms. Indeed, more than 100,000 venomous species are distributed among different taxa. The poison contains what we refer by toxins, but venoms are essentially a mixture of many compounds including proteins, peptides, salts, organic molecules, amino acids, and neurotransmitters-like molecules that produce a synergically toxic effect. In general, the mechanism of action involves hydrolytic enzymes that degrade tissues, allowing other toxins to diffuse up on their targets mainly in the nervous or cardiovascular system. Among these targets we can highlight, membrane receptors, ion channels and enzymes that regulate the metabolism of excitable cells. Such toxins usually act at very low concentrations on their targets, causing a drastic change in important physiological functions that eventually lead to death.

Toxins are widely distributed among metazoans and there are some venomous lineages both in vertebrates and invertebrates. Within vertebrates, snakes represent one of the major sources of toxins, and have been so far studied due to its powerful toxins and biomedical interest. By contrast, Cnidarians, which are grouped in the largest phylum of venomous animals, remain still unexplored. The species of the phylum Cnidaria commonly possess specialized stinging cells called nematocyst that produce and inject into prey or predator a mixture of toxins, whilst snakes possess maxillary venom glands coupled to front or rear fangs. Many toxins like enzymes, protease inhibitors, ion channels modulators, have been isolated and characterized from both groups. Venoms often contain a group of peptide/protein toxins with neurotoxic and cardiotoxic activities. However, Cuban and Portuguese cnidarians represent a rich source of toxins but remain mostly underexplored. Similarly, there are no studies addressing the production of toxins in snakes from Cuba, even though clinical symptoms have been reported after bites of some colubrids.

The main goal of this project is to perform the proteomic characterization of toxins from Cuban and Portuguese cnidarians, and to profile the Harderian gland transcriptome from Cuban snakes. The generated information will increase the information about such toxins and its protein-encoding genes. Moreover, the characterization of novel toxins may allow us to discover novel cell excitability modulators as a source of new pharmacological tools or therapeutic products. In addition, the new findings will provide insight into the evolutionary history of the molecular diversification of toxins and its venom-encoding genes.

The phylum Cnidaria is an ancient group of venomous animals, specialized in the production and delivery of toxins. Many species belonging to the class Anthozoa have been

studied and their venoms often contain a group of peptides of less than 10 kDa that act upon ion channels. These peptides and their targets interact with high affinity producing neurotoxic and cardiotoxic effects, and even death, depending on the dose and the administration pathway. Zoanthinaria (Cnidaria) is an order of the Subclass Hexacorallia, class Anthozoa, and unlike sea anemone (order Actiniaria), neither its diversity of toxins nor the *in vivo* effects of the venoms has been exhaustively explored. Unlike sea anemones, proteomics studies aiming toxins discovering from the order Zoanthidea are scarce. There are only few reports about the toxicological properties of its members and their toxins composition is scarce. In **CHAPTER 2**, some toxicological tests on mice with a low molecular weight fraction obtained by gel filtration in Sephadex G-50 from *Zoanthus sociatus* crude extract were assessed. The toxicological effects of the studied fraction seem to be mostly autonomic and cardiotoxic, causing death in a dose dependent manner with a LD₅₀ of 792 µg/kg. Moreover, at a sub-lethal dose the active fraction accelerated the KCl-induced lethality in mice.

Information obtained in the **CHAPTER 2** shed light about the molecular mass composition of the fraction from *Z. sociatus*, which resulted lethal to mice. However, the identification and nature of the components of such fraction remains unknown. Therefore, in **CHAPTER 3**, a mass spectrometry analysis of a low molecular weight (LMW) fraction previously reported as lethal to mice was performed. The low molecular weight (LMW) fraction was obtained from the *Z. sociatus* by crude extract gel filtration in Sephadex G-50. Subsequently, the fraction of interest was characterized by mass spectrometry analyses. However, no sea anemones-like toxins were identified rather than microcystin masses. Subsequent reversed-phase C18 HPLC (in isocratic elution mode) and mass spectrometry analyses corroborated the presence of the cyanotoxin Microcystin-LR (MC-LR). To the best of our knowledge, this finding constitutes the first report of MC-LR in *Z. sociatus*, and one of the few evidences of such cyanotoxin in cnidarians.

Currently around 250 toxic compounds from cnidarians have been identified including peptides, proteins, enzymes, protease inhibitors and non-proteinaceous substances. Unexpectedly, no cnidarian toxin was identified into the components of the low molecular weight fraction from *Z. sociatus*. In those cases, (**CHAPTER 2** and **CHAPTER 3**), more classical methods were applied based on purification and toxicological tests of semi purified fractions. However, until now, most of the toxins from cnidarians belonging to sea anemones were discovered by classical purification approaches combined with guided-bioassays protocols. Recently, the use of high-throughput methodologies increased significantly the number of proteins and toxins identified, but mostly in other groups of venomous animals. Portugal has a diverse representation of sea anemones, which are a promising source of bioactive compounds. Despite some of them are intertidal species and provide relative easy

access, the knowledge about its toxins production is still limited. Thus, in **CHAPTER 4**, shotgun proteomics of the whole-body extract from the unexplored sea anemone *Bunodactis verrucosa* was profiled. The proteomic analyses applied were based on two-dimensional gel electrophoresis combined with MALDI-TOF/TOF and gel-free approaches carried out by nano-LC coupled to a hybrid Ion trap mass spectrometer (LTQ Orbitrap). In total, 413 proteins were identified by shotgun proteomics approaches. The Kyoto Encyclopedia of Genes and Genomes analyses (KEEG) obtained from the Blast2Go software, revealed that the most represented enzymatic pathways were Purine metabolism, Thiamine metabolism, Biosynthesis of antibiotics and Glycolysis/Gluconeogenesis. Moreover, some toxins including metalloproteinases and neurotoxins were successfully identified. The mechanism of action of such toxins in prey catching and feeding is proposed, which seemingly act synergically. The present work provides the first map of the proteome of the sea anemone *B. verrucosa*.

While the previous investigations have characterized cnidarians toxins, snakes represent also an interesting target as toxins source. Unlike cnidarian toxins, snakes' venoms have evolved in accelerated manner generating a wide variety of toxins. Species which have been isolated for long periods of time, like the cuban snakes, are more likely to develop new genetic strategies, which can result in biological novelties even in front fanged snakes and colubrids. However, most of the current toxinologists address their effort to characterized the venomous repertoire in snakes with medical relevance. Integrated "omics" profiling venom glands are growing up, but just a few studies have been performed in the Colubrid family. Within them, the transcriptomic analyses of the Duvernoy's analogues showed similarities in toxin transcripts composition to Viperidae and Elapidae snakes.

Nonetheless, there is another gland in colubrids called the Harderian gland, which is relative larger in some species and is anatomically connected to the vomeronasal organ (VNO) via the nasolacrimal duct. The function of this gland's secretion remains unknown, but have been proposed to play a role in several functions as: a source of saliva, pheromones, thermoregulatory lipids and growth factors; part of a retinal-pineal axis, as a photoprotective organ, a site of immune response and osmoregulation. However, the amount of venom produced by some species constitute a limiting factor. Next generation sequencing (NGS) approaches has become an invaluable tool to characterize several tissues including glands from the snake's head. Thus, a transcriptomic profile of three species of colubrids from Cuba was carried out in **CHAPTER 5**. Herein, the Harderian gland transcriptome of three snakes from Cuba was profiled: *Caraiba andreae* (Ca), *Cubophis cantherigerus* (Cc) and *Tretanorhinus variabilis* (Tv), based on Illumina HiSeq2000 100 bp paired-end. Apart from some housekeeping genes related to ribosomal and cellular components, the most expressed contigs of the Harderian gland were related to transport/binding and snake's

toxins. Indeed, most known classes of the snake's toxins were identified. Therefore, the Harderian gland could be deeply involved in binding/transport, maybe related to vomerolfaction, but also with toxins production that could be addressed to protection against microorganism, perhaps even to kill preys.

In general, most of the obtained results are novelties, constituting first reports. Much of the value of the developed work and its related discoveries is given by the approaches applied. Such methodologies cover protocols from classical approaches to more recent like Next Generation Sequencing aiming protein, toxins and genes characterization in two main groups of venomous animals. In cnidarians, whole body-extract preparation, fractionation techniques, in vivo toxicological test, and both gel-based and gel-free spectrometry analysis were employed, while transcriptomic approaches were applied in snakes. The combination of analytical techniques allowed the identification of non-proteinaceous and proteinaceous components in two cnidarians species. In the case of *Z. sociatus*, a low molecular weight fraction resulted lethal to mice, but no related cnidarians toxin was identified. Unexpected, MC-LR and other cyanotoxins masses were detected. The identification of MC-LR constitutes the first report for this species and one of the few for the Phylum cnidarian. In the case of the Portuguese sea anemone *B. verrucosa*, it was reported here by the first time the complete proteome map, including some toxins.

On the other hand, the Harderian gland transcriptome was obtained from three Cuban colubrids, giving insights of such gland function. In addition, this result constitutes the first transcriptome of the Harderian gland in reptiles, and the second in vertebrates. Despite this gland has never been associated with venom function before, some toxins occurred among the most expressed transcripts. Although High-throughput analysis of both shotgun proteomics and transcriptomics resulted especially suitable in this study, the classical methods are still needed. Altogether, high-throughput approaches combined with classical bioassays-guided chromatographic purifications, provided an integrated information for protein/toxins characterization.

Keywords

biological activity; toxins; venoms; *Zoanthus sociatus*; Zoanthidea; *Bunodactis verrucosa*; sea anemones; Anthozoa; Cnidaria; LD₅₀ mice; proteomics; transcriptomics; Microcystins; MC-LR; Sephadex G50; RP-HPLC; MALDI-TOF/TOF; shotgun proteomics; proteins; Two-dimensional gel electrophoresis; lipocalin; binding; vomerolfaction; defense; *Caraiba andreae*; *Cubophis cantherigerus*; *Tretanorhinus variabilis*; Dipsadinae; Cuba; Portugal

Resumo

A presença de toxinas é uma característica que confere vantagens significativas aos animais venenosos na luta pela sobrevivência. Ao longo da evolução, muitos grupos de animais desenvolveram de forma independente tecidos especializados acoplados a um sistema de descarga como agulhas, arpões, para produzir e inocular venenos. Na verdade, mais de 100.000 espécies venenosas são distribuídas entre diferentes taxa. O veneno contém o que habitualmente referimos por toxinas, mas os venenos são essencialmente uma mistura de muitos compostos, incluindo proteínas, péptidos, sais, moléculas orgânicas, aminoácidos e moléculas semelhantes a neurotransmissores que produzem um efeito sinergicamente tóxico. Em geral, o mecanismo de acção envolve enzimas hidrolíticas que degradam os tecidos, permitindo que outras toxinas se difundam em seus alvos principalmente no sistema nervoso ou cardiovascular. Entre esses alvos podemos destacar, receptores de membrana, canais iónicos e enzimas que regulam o metabolismo de células excitáveis. Tais toxinas geralmente agem em concentrações muito baixas em seus alvos, causando uma mudança drástica em importantes funções fisiológicas que eventualmente levam à morte.

As toxinas encontram-se amplamente distribuídas entre metazoários e existem linhagens venenosas tanto em vertebrados quanto em invertebrados. Dentro dos vertebrados, as cobras representam uma das principais fontes de toxinas, e até agora foram estudadas devido às suas poderosas toxinas e ao seu interesse biomédico. Em contraste, os cnidários, que estão agrupados no maior filo de animais venenosos, permanecem ainda inexplorados. As espécies do filo Cnidaria geralmente possuem células urticantes especializadas chamadas nematocistos que produzem e injectam em presas ou predadores uma mistura de toxinas, enquanto as cobras possuem glândulas de veneno maxilar acopladas às presas dianteiras ou traseiras. Muitas toxinas, como enzimas, inibidores de protease, moduladores de canais iónicos, foram isoladas e caracterizadas para ambos os grupos. Os venenos geralmente contêm um grupo de toxinas de péptidos/proteínas com actividades neurotóxicas e cardiotoxinas. No entanto, os cnidários cubanos e portugueses representam uma rica fonte de toxinas, mas permanecem principalmente subexplorados. Da mesma forma, não há estudos que abordem a produção de toxinas em cobras de Cuba, mesmo quando sintomas clínicos tenham sido relatados após mordidas dessas cobras.

O objectivo principal deste projecto é realizar a caracterização proteómica de toxinas de cnidarians cubanos e portugueses e perfilar o transcriptoma da glândula de Harder de cobras cubanas. A informação gerada aumentará o conhecimento sobre tais toxinas e seus genes que codificam proteínas. Além disso, a caracterização de novas toxinas pode permitir descobrir novos moduladores de excitabilidade celular como fonte de novas ferramentas

farmacológicas ou produtos terapêuticos. Além disso, as novas descobertas fornecerão uma visão da história evolutiva da diversificação molecular das toxinas e seus genes codificadores de veneno.

O filo Cnidaria é um antigo grupo de animais venenosos, especializado na produção e descarga de toxinas. Muitas espécies pertencentes à classe Anthozoa foram estudadas e seus venenos geralmente contêm um grupo de péptidos de menos de 10 kDa que actuam sobre os canais iónicos. Estes péptidos e seus alvos interagem com efeitos neurotóxicos e cardiotoxicos de alta afinidade, e até mesmo a morte, dependendo da dose e da via de administração. Zoanthinaria (Cnidaria) é uma ordem da Subclasse Hexacorallia, classe Anthozoa, e ao contrário da anémona do mar (ordem Actiniaria), nem a sua diversidade de toxinas nem os efeitos *in vivo* dos venenos foram explorados exaustivamente. Ao contrário das anémonas do mar, os estudos de proteómica visando a descoberta de toxinas na ordem Zoanthidea são escassos. Há apenas alguns relatórios sobre as propriedades toxicológicas de seus membros e a composição de toxinas conhecidas é escassa. No **CAPÍTULO 2**, foram avaliados alguns testes toxicológicos em ratos com fracção de baixo peso molecular obtidos por filtração em gel em Sephadex G-50 a partir do extracto bruto *Zoanthus sociatus*. Os efeitos toxicológicos da fracção estudada parecem ser principalmente cardiotoxicos, causando a morte de maneira dependente da dose com um DL50 de 792 µg/kg. Além disso, em uma dose sub-letal, a fracção activa acelerou a letalidade induzida por KCl em ratinhos.

As informações obtidas no **CAPÍTULO 2** revelam a composição da massa molecular da fracção de *Z. sociatus*, que resultou letal para ratos. No entanto, a identificação e a natureza dos componentes dessa fracção permanecem desconhecidas. Assim, no **CAPÍTULO 3**, realizou-se uma análise de espectrometria de massa de uma fracção de baixo peso molecular (LMW) anteriormente relatada como letal para ratos. A fracção de baixo peso molecular (LMW) foi obtida a partir do *Z. sociatus* por filtração de gel de extracto bruto em Sephadex G-50. Posteriormente, a fracção de interesse caracterizou-se por análises de espectrometria de massa. No entanto, não foram identificadas toxinas de anémonas do mar em vez disso, detectaram-se massas de microcistina. O HPLC subsequente de fase reversa C18 e as análises de espectrometria de massa corroboraram a presença da cianotoxina Microcistina-LR (MC-LR). No nosso melhor conhecimento, essa descoberta constitui a primeira referência de MC-LR em *Z. sociatus* e uma das poucas evidências dessa cianotoxina em cnidários.

Actualmente, cerca de 250 compostos tóxicos de cnidários foram identificados, incluindo péptidos, proteínas, enzimas, inibidores de proteases e substâncias não proteináceas. Inesperadamente, nenhuma toxina de cnidários foi identificada nos componentes da fracção de baixo peso molecular de *Z. sociatus*. Nesses casos,

(**CAPÍTULO 2** e **CAPÍTULO 3**), métodos mais clássicos foram aplicados com base em testes de purificação e métodos toxicológicos de fracções semi-purificadas. No entanto, até agora, a maioria das toxinas de cnidários pertencentes a anémonas de mar foram descobertas por abordagens clássicas de purificação combinadas com protocolos de bioensaios guiados. Recentemente, o uso de metodologias de alto rendimento aumentou significativamente o número de proteínas e toxinas identificadas, mas principalmente em outros grupos de animais venenosos. Portugal tem uma representação diversificada das anémonas do mar, que são uma fonte promissora de compostos bioativos. Apesar de algumas delas serem espécies intertidais e proporcionarem um acesso relativamente fácil, o conhecimento sobre a produção de toxinas ainda é limitado. Assim, no **CAPÍTULO 4**, a proteómica do extracto total da anémona do mar *Bunodactis verrucosa* foi efectuada. As análises proteómicas aplicadas foram baseadas em electroforese em gel bidimensional combinada com abordagens MALDI-TOF/TOF e isentas de gel realizadas por nano-LC acopladas a um espectrómetro de massa híbrido (LTQ Orbitrap). No total, 413 proteínas foram identificadas por abordagens proteómicas. As análises da Enciclopédia de Quioto de Genes e Genomas (KEEG) obtidas do software Blast2Go, revelaram que as vias enzimáticas mais representadas foram o metabolismo de purina, o metabolismo de tiamina, a bio-síntese de antibióticos e a glicólise/gluconeogénese. Além disso, algumas toxinas incluindo metaloproteinases e neurotoxinas foram identificadas com sucesso. Foi proposto o mecanismo de acção de tais toxinas em presas, captura e alimentação, que aparentemente agem de forma sinérgica. O presente trabalho fornece o primeiro mapa do proteoma da anémona do mar *B. verrucosa*.

Enquanto as investigações anteriores caracterizaram toxinas de cnidários, as cobras representam também um alvo interessante como fonte de toxinas. Ao contrário das toxinas de cnidários, os venenos de cobras evoluíram de forma acelerada gerando uma grande variedade de toxinas. As espécies que foram isoladas por longos períodos de tempo, como as cobras cubanas, são mais propensas a desenvolver novas estratégias genéticas, o que pode resultar em novidades biológicas. No entanto, a maioria dos ensaios toxicológicos actuais abordam o repertório venenoso em cobras com relevância médica. As tecnologias "omics" integradas que perfilam glândulas de veneno tem sido usadas mais recentemente, mas apenas alguns estudos foram realizados na família Colubridea. As análises transcriptómicas dos análogos da glândula de Duvernoy mostraram semelhanças na composição dos transcritos de toxinas para as cobras Viperidae e Elapidae.

No entanto, há outra glândula em colubrideos chamada de glândula de Harder, que é relativamente maior em algumas espécies e está anatomicamente ligada ao órgão vomeronasal (VNO) através do ducto nasolacrimal. A função da secreção desta glândula permanece desconhecida, mas foi proposto desempenhar um papel em várias funções

como: uma fonte de saliva, feromonas, lípidos termo-reguladores e factores de crescimento; parte de um eixo retinal-pineal, como órgão fotoprotector, local de resposta imune e osmoregulação. No entanto, a quantidade de veneno produzida por algumas espécies constitui um factor limitante. As abordagens de sequenciação de última geração (NGS) tornaram-se uma ferramenta inestimável para caracterizar vários tecidos, incluindo glândulas de cobras. Assim, um perfil transcriptómico de três espécies de colubrídeos de Cuba foi realizado no **CAPÍTULO 5**. Aqui, foi descrito o transcriptoma da glândula de Harder de três cobras de Cuba: *Caraiba andreae* (Ca), *Cubophis cantherigerus* (Cc) e *Tretanorhinus variabilis* (Tv), com base em Illumina HiSeq2000. Além de alguns genes relacionados com os componentes ribossómicos e celulares, os *contigs* mais expressos da glândula *Harderian* estavam relacionados ao transporte/ligação de toxinas de cobra. Na verdade, as classes mais conhecidas de toxinas de cobras foram identificadas. Portanto, a glândula *Harderian* poderá estar envolvida na olfacção vomeronasal, mas também com a produção de toxinas eventualmente relacionadas com protecção contra o microrganismos, possibilitando eventualmente até para matar presas.

Em geral, a maioria dos resultados obtidos são de destacar. Grande parte do valor do trabalho desenvolvido e suas descobertas resulta das novas abordagens aplicadas. Tais metodologias abrangem desde os protocolos mais clássicos aos mais recentes, como a sequenciação de última geração, visando a caracterização de proteínas, toxinas e de genes em dois grupos principais de animais venenosos. Em cnidários, foram estudadas preparações de extractos completos, técnicas de fraccionamento, testes toxicológicos *in vivo* e análises de espectrometria, enquanto que as abordagens transcriptómicas foram aplicadas em cobras. A combinação de técnicas analíticas permitiu a identificação de componentes não proteináceos e proteináceos em duas espécies de cnidários. No caso de *Z. sociatus*, uma fracção de baixo peso molecular resultou letal para ratinhos, mas nenhuma toxina de cnidários relacionada foi identificada. Inesperadamente, MC-LR e outras massas de cianotoxinas foram detectadas. A identificação do MC-LR constitui o primeiro relatório para esta espécie e um dos poucos para o este Filo. No caso da anémone marinha portuguesa *B. verrucosa*, foi relatado aqui pela primeira vez o mapa do proteoma completo, incluindo algumas toxinas.

Por outro lado, o transcriptoma da glândula de Harder foi obtido de três colubrídeos cubanos, dando informações sobre a função dessa glândula. Além disso, esse resultado constitui o primeiro transcriptoma da glândula de Harder em répteis e o segundo em vertebrados. Apesar desta glândula nunca ter sido associada com a função de veneno antes, algumas das toxinas detectadas foram muito expressas. Embora a análise do proteoma e do transcriptoma resulte especialmente adequada neste estudo, os métodos clássicos ainda são necessários. No total, abordagens de alto rendimento combinadas com

purificações cromatográficas orientadas por bioensaios clássicos, forneceram uma informação integrada para caracterização de proteínas/toxinas.

Palavras-chave

Actividade biológica; toxinas; venenos; *Zoanthus sociatus*; Zoanthidea; *Bunodactis verrucosa*; anémonas do mar; Anthozoa; Cnidaria; LD₅₀ ratinho; proteómica; transcriptómica; Microcystins; MC-LR; Sephadex G50; RP-HPLC; MALDI-TOF/TOF; *shotgun proteomics*; proteínas; gel de eletroforese a duas-dimensões; *lipocalin*; ligação; olfacção vomeronasal; defesa; *Caraiba andreae*; *Cubophis cantherigerus*; *Tretanorhinus variabilis*; Dipsadinae; Cuba; Portugal

Table of Contents

Acknowledgements-----	iv
Abstract -----	vi
Keywords -----	ix
Resumo -----	x
Palavras-chave-----	xiv
Table of Contents-----	xvi
List of Tables-----	xx
List of Figures-----	xxii
List of Abbreviations-----	xxiv
Chapter 1 -----	1
Introduction-----	1
1.1 General Introduction-----	3
1.2 Diversity and Physiological Effects of toxins -----	4
Toxins as modulator of ion channels-----	5
Cytolytic toxins and hydrolytic enzymes-----	8
Protease Inhibitors-----	8
Toxins targeting physiometabolic processes-----	9
Immune response against animal toxins -----	11
1.3 Biomedical potential of toxins -----	12
1.4 General methods in the discovering of new toxins-----	13
1.4.1 Isolation and characterization of cnidarians toxins -----	13
General protocols on cnidarian venom extraction-----	13
General procedures in the purification of cnidarians toxins-----	15
Proteomics in the study of venom-----	16
1.4.2 Isolation and characterization of cnidarians toxins -----	18
Sequencing Platforms in NGS -----	18
Transcriptomic approaches-----	19
Transcriptome advances in the study of venom snake glands-----	20
1.5 Animals models in the characterization of toxins-----	21
Overview of toxicological tests-----	21
Behavioral assessment in pharmacological studies -----	22
1.6 General aims-----	23
Cnidarians -----	23
Snakes -----	24

1.7 Thesis outline -----	25
Chapter 2 -----	27
Insights into the Toxicological Properties of a Low Molecular Weight Fraction from <i>Zoanthus sociatus</i> (Cnidaria)-----	27
2.1 Abstract -----	29
2.2 Introduction -----	31
2.3 Results and Discussion-----	32
2.4 Experimental Section -----	35
2.5 Conclusion-----	36
2.6 Supplementary Information-----	37
Chapter 3 -----	39
Microcystin-LR Detected in a Low Molecular Weight Fraction from a Crude Extract of <i>Zoanthus sociatus</i> -----	39
3.1 Abstract -----	41
3.2 Introduction -----	43
3.3 Results and Discussion-----	44
3.4 Conclusions -----	53
3.5 Materials and Methods -----	53
3.6 Supplementary Materials-----	56
Chapter 4 -----	57
Proteomic analyses of the unexplored sea anemone <i>Bunodactis verrucosa</i> -----	57
4.1 Abstract -----	61
4.2 Introduction -----	63
4.3 Materials and Methods -----	64
4.4 Results and Discussion-----	67
4.5 Conclusion-----	79
4.6 Supplementary Materials-----	80
Chapter 5 -----	81
Venomous repertoire involved in the Harderian gland transcriptomes of three snakes (Colubridae) from Cuba -----	81
5.1 Abstract -----	85
5.2 Background-----	87
5.3 Methods -----	89
5.4 Results and Discussion-----	94
5.5 Conclusions -----	109
5.6 Supplementary Materials-----	109
Chapter 6 -----	111
General Discussion-----	111

General Discussion-----	113
Chapter 7 -----	121
Conclusions and Future Perspectives -----	121
Concluding Remarks-----	123
Future Perspectives -----	124
General Conclusions-----	125
Chapter 8 -----	127
References -----	127
References -----	129

List of Tables

Table 4.1: Blast Search summary.	69
Table 4.2. Top twenty KEGG pathways.....	75
Table 4.3. Potential toxins.	77
Table 5.1. Statistics of the transcriptomes.....	95
Table 5.2. Description of the 25 most expressed contigs.	98
Table 5.3. Relative expression of snake Harderian gland toxins.	107

List of Figures

Figure 1.1. Transitions between ion channel states.....	6
Figure 1.2. Toxins as modulator of site 3 and 4 of the sodium (Na ⁺) voltage-dependent ion channel.	7
Figure 1.3. Animal toxins affect insulin secretion.....	10
Figure 2.1. Sephadex G50 gel filtration chromatogram of <i>Z. sociatus</i> crude extract.	32
Figure 2.2. Acute toxicity assay.....	34
Figure 2.3. The low molecular weight fraction ZsG50-III accelerated the KCl-induced time to cardiac arrest.	34
Figure 3.1. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF) mass spectrum of the Sephadex G-50 fraction called ZsG50-III.	45
Figure 3.2. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass-spectra of two peaks of interest from fraction ZsG50-III.....	46
Figure 3.3. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) analysis of two signals from fraction ZsG50-III mass spectrum.	46
Figure 3.4. The figure shows absolute intensity (a.i) versus mass-to-charge ratio (m/z) in the m/z range 960–1050 from the MS analysis of fraction ZsG50-III.	48
Figure 3.5. Analytical profile of fraction ZsG50-III, obtained by RP-HPLC.	48
Figure 3.6. Chromatogram of fraction ZsG50-III subjected to reversed-phase C18 HPLC...	49
Figure 3.7. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrum of the two RP-HPLC peaks.	50
Figure 3.8. MS/MS of MC-LR.	51
Figure 4.1. Two-dimensional gel electrophoresis and identification of soluble proteins from <i>Bunodactis verrucosa</i>	68
Figure 4.2. Blast2Go Data Distribution chart.	72
Figure 4.3. Blast2Go Species distribution chart.....	73
Figure 4.4. Blast2Go hits Gene Ontology (GO) annotation.....	74
Figure 5.1. Two cephalic glands in <i>Caraiba andreae</i> , the Harderian gland (Hg) and the Duvernoy's gland (Dg).	88
Figure 5.2. Most expressed contigs.....	97
Figure 5.3. Gene ontology (GO) statistics of non-toxin contigs.....	102
Figure 5.4. Gene ontology (GO) annotation of non-toxin contigs by level 2.	103
Figure 5.5. Gene ontology (GO) annotation of non-toxin contigs by level 2.	104
Figure 5.6. Gene ontology (GO) annotation of non-characterized proteins by level 2.	105
Figure 5.7. Relative expression of snake's toxins detected in the transcriptome of the Cuban colubrids	106

List of Abbreviations

2DE	Two-dimensional gel electrophoresis
8-CRS	eight Costa Rican snakes
a.i	Absolute intensity
Adda	3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-deca-4,6-dienoic acid
AT	Adenin-Timin
AT (%)	The AT percentage in sequence reads
AVE	average
<i>B. verrucosa</i>	<i>Bunodactis verrucosa</i>
BP	Biological Process
BPP	Bradykinin potentiating and C-type natriuretic peptides
Ca	<i>Caraiba andreae</i>
CC	Cellular Components
CHAPS	Surfactant used in the laboratory to solubilize biological macromolecules such as proteins (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate).
CREGF	Cysteine-rich with EGF-like domain
CRISP	Cysteine-rich secretory protein
CTL	C-type lectin
Cc	<i>Cubophis cantherigerus</i>
DTT	Dithiothreitol
FA	Formic acid
FASP	Filter-aided sample preparation
FDR	False Discovery Rate
fwhm	Full width at half maximum
Gb	Gigabase
GC	Guanine-Cytosin
GC (%)	The GC percentage in sequence reads
GO	Gene ontology
HPLC	High Performance Liquid Chromatographic
HYAL	Hyaluronidase
IEF	Isoelectric focusing (electrofocusing)
IF	insoluble protein fraction

int	Peak's corrected intensity
KEEG	Kyoto Encyclopedia of Genes and Genomes
KUN	Kunitz-type protease inhibitor
L50	number of contigs whose length is $\geq$ N50
LAO	L amino-acid oxidase
LC/ESI-Q-ToF	Liquid Chromatography/Electrospray ionization coupled to Quadrupole Time-of-flight
LC-MS	Liquid Chromatography–Mass Spectrometry
LD	Lethal Dose
LMW	Low molecular weight
m/z	Mass-to-charge ratio
MALDI-TOF/TOF	Matrix assisted laser desorption/ionization time-of-fly/time-of-fly
mass	Neutral mass
MC-LR	Microcystin LR
MEC	most expressed contigs
MF	Molecular Function
Molecular Function	MF
MS	Mass Spectrometry
MS	Mass spectra
mtDNA	Mitochondrial DNA
MYO	Myotoxin (crotonamine)
N50	average contigs length
NF	Neurotrophic factor
NGF	Nerve growth factor
NGS	New generation sequencing
NMWL	nominal molecular weight limit
NORINE	Non-ribosomal peptide database
nt	Nucleotide
NUC	Nucleotidase
<i>P. caribaeorum</i>	<i>Palythoa caribaeorum</i>
PDA	Photodiode Array
PDE	Phosphodiesterase
PDI	Protein disulfide isomerase
PE	paired-end reads

PLA2	Phospholipase A2
Q	$Q = -10 \log_{10}(\text{error rate})$
Q20(%)	The percentage of bases in which the phred score is above 20
Q30(%)	The percentage of bases in which the phred score is above 30
r.int	Relative intensity
resol	resolution
RIN	RNA Integrity Number
RP	Reverse Phase
RPKM	Reads Per Kilobase per Million mapped reads
RPKM-MAX	Maximum Value of RPKM
RPKM-MIN	Minimum value of RPKM
S/N	Peak's signal-to-noise ratio
SB	Buffer SB is a buffer solution made up of sodium borate, usually 1–10 mM at pH 8.0. used in agarose and polyacrylamide gel electrophoresis
SDS-PAGE	Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis
SF	soluble protein fraction
stats	statistics
SVMP	Snake venom metalloproteinase
SVSP	Snake venom serine proteinase
TFA	Trifluoroacetic acid
TFA	Trifluoroacetic acid
TPM	transcripts per million
Tv	<i>Tretanorhinus variabilis</i>
TTX	Tetrodotoxin
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
VESP	Vespryn (ohanin-like)
VF	Venom factor; WAP
VNO	Vomeronasal organ
WAP	Waprin
<i>Z. sociatus</i>	<i>Zoanthus sociatus</i>

CHAPTER 1

Introduction

1.1 General Introduction

Throughout evolution many group of animals have been developing strategies to survive, and one of the most successful adaptive mechanism is the production of venoms. Indeed, more than 100,000 venomous species are distributed among different taxa, such as chordates (fishes, amphibians, reptiles, mammals), echinoderms (starfishes, sea urchins), mollusks (cone snails, octopus), annelids (leeches), nemertines, arthropods (arachnids, insects, myriapods) and cnidarians (sea anemones, jellyfish, corals) [1, 2]. These venoms commonly contain a mixture of many compounds including proteins, peptides, salts, organic molecules such as polyamines, amino acids, and neurotransmitters [3-7], produced by exocrine glandules coupled to a delivery system like fangs, needles, harpoons [8]. The component of the venoms, commonly known as toxins, may act synergically upon different types of target within cells and on the plasma, such as ion channels, enzymes [1], causing the alteration of important physiological processes [9-12]. Moreover, venoms also contain protease inhibitors and stabilizing agents to prevent degradation against external and internal biotic and abiotic factors, like proteases and temperature [1]. Thus, the venoms can be preserved into the gland for weeks. Hence, these features confer significant advantages to those venomous animals to obtain food and/or avoiding predators.

Large genes families encoding for diverse peptide/protein toxins are common in poisonous animals [13-15]. The origin of toxins families can be explained by gene recruitment events in which an ordinary protein-coding gene is duplicated (likely involved in a key regulatory process) [2]. Then, the new gene can be expressed in a specialized tissue [2], but eventually could be subsequently duplicated resulting in multigene families, thus generating several neofunctionalizations [4, 16]. The deletion of some copies, or degradation of non-functional copies or pseudogenes could then happened [17]. Such toxins share several conserved features, such as disulfide cross-link and stable molecular scaffolds [2], as well as receptor binding sites [18]. The presence of cysteines produces cross-linking disulfide bonds conferring stabilization to the protein structure, giving also protection against high temperature. The conserved signatures of the toxin sequences are related to their physiological targets [1]. For example, the 'KY-dyad' (Lys-Tyr) was found as the most critical feature for the biding to potassium channels [19, 20], even when the molecular scaffold between them is different [14]. In addition, other remarkable pattern is the 'RGD' motifs (Arg-Gly-Asp), which is involve in the interaction to integrins [21]. However, these conserved sequences have been replaced by other tripeptide in the same position [22], which suggests that a variety of combinations could play a similar role.

Several gene-family members in snakes, cone snails and scorpions have been evolving in an accelerated manner compared with other regions of the genome. Hence,

indicating that these genes have been evolving under positive Darwinian selection (also known by diversifying selection) [23-25], while sea anemone neurotoxins-encoding genes exhibit extreme conservation at the nucleotide level [14]. These later genes have evolved mainly via concerted evolution, although some genes evolved under diversifying selection, suggesting that both mechanism of evolution may occur simultaneously [26]. However, many of the features mentioned are also shared by non-toxic peptides, proteins or domains that do not belong to the venom secretion [1]. This justifies the needed of integrating multiple approaches like proteomics, genomics, transcriptomics, evolutionary-based analyses, and even guided bioassays to avoid false positive detections in the discovering of new toxins. High-throughput analyses combined with integrated based-evolutionary approaches provide a suitable platform to identify animals toxins [1], but still represents a challenge in the identification whether new configurations or substitutions occurs.

1.2 Diversity and Physiological Effects of toxins

In general, the term “toxin” has been ascribed to almost any compound capable of producing significant change in at least one physiological process of the victim. However, the terms “venom” or “animal toxin” are associated with a secretion produced in a specialized tissue capable of causing physiological disorder, eventually death. Indeed, the production of toxins constitutes a great advantage, since they can powerfully act in a low concentration, mainly upon membrane receptors, ion channels and enzymes that regulate the metabolism of excitable cells [2, 9, 14, 27]. Among them, it could be mentioned the AVIT/colipase/prokineticin, CAP, chitinase, cystatin, defensins, hyaluronidase, Kunitz, lectin, lipocalin, natriuretic peptide, peptidase S1, phospholipase A₂, sphingomyelinase D, and SPRY [2], which are present in well-known venomous groups from cnidarians to mammals. However, the presence of such toxins, have been also found in fleas, leeches, kissing bugs, mosquitoes, ticks and mammals [2].

Toxins show also ecological benefits, and as previously mentioned its function play important environmental role in predation and defense. As an example, some neurotoxin from sea anemones exhibit more specificity against sodium (Na⁺) voltage-dependent ion channels of crustaceans (arthropods in generals) than mammals [28]. This is likely related with the evolution process which pressure toxins to be more effective against animals that share the same habitat, as potential prey or predators. Similar features are found in scorpions and arachnids’ toxins, where insect constitutes a relevant source of food. These features constitute a hint for future selective insecticides [18, 29, 30]. Other important ecological value of toxins is revealed in insectivorous mammals, which used toxins to subdue larger preys depending of doses and administration pathways. At least in insects

and small mammals, toxins can produce death or just induce a vegetative state in the prey, allowing the preservation of considerable amount of food for long time [31]. Parasitoid wasp like the tarantula hawk *Hemipepsis ustulata* use similar mechanism to neutralize tarantulas. Wasps inject a non-lethal venom into the prey inducing permanent paralyzes, but keeping it alive [32], allowing their larvae to feed on the host after hatching as long as possible for freshness [33].

Toxins as modulator of ion channels

Most of the toxins have ion channels as targets [34, 35], because they are involved in many critical physiological functions [36-38], like excitability in neurons [39-43] or cardiomyocytes [44-47]. Ion channels are dynamic transmembrane proteins which modify the permeability of the membrane through the conductance to ions like Na^+ , K^+ , Ca^{2+} , Cl^- [48]. Toxin can mimic the natural ligands, hormones, neurotransmitters or cellular messengers. Besides, they are ion selective and can block permanent or partially such ion channels, modulating the conformation of the protein, thus altering the ion flow. The most classical states that characterize the ion channels are closed, open and inactivated (Figure 1.1). Thus, they fluctuate between closed and open states, but closed state predominates on resting state rather than the open state, which is a fleeting event [49]. Ion channels are also stimulus-specific, (e.g. transmembrane voltage, temperature, ligands, pH) hence the nature and magnitude of the stimulus can affect the probability of remaining more time in one state, gather or not the ion flow into the channel pore [50, 51].

Cnidarians, cone snails, scorpions, spiders, insects and snakes constitute a great source of toxins as ion channel modulators. Among the famous are dendrotoxins [52, 53], which are small proteins containing 57-60 amino acid residues cross-linked by three disulphide bridges, isolated from mamba (*Dendroaspis*) snakes [54]. The α -dendrotoxin from green mamba *Dendroaspis angusticeps* and toxin I from the black mamba *Dendroaspis polylepis* block a variety of potassium channel in the low nanomolar range [54]. Some dendrotoxins homologues, like α -, β -, γ - and δ -dendrotoxins occur in the same venom [55], showing homology in the sequences and folding with Kunitz-type serine protease inhibitors, but without significant anti-protease activity [56, 57].

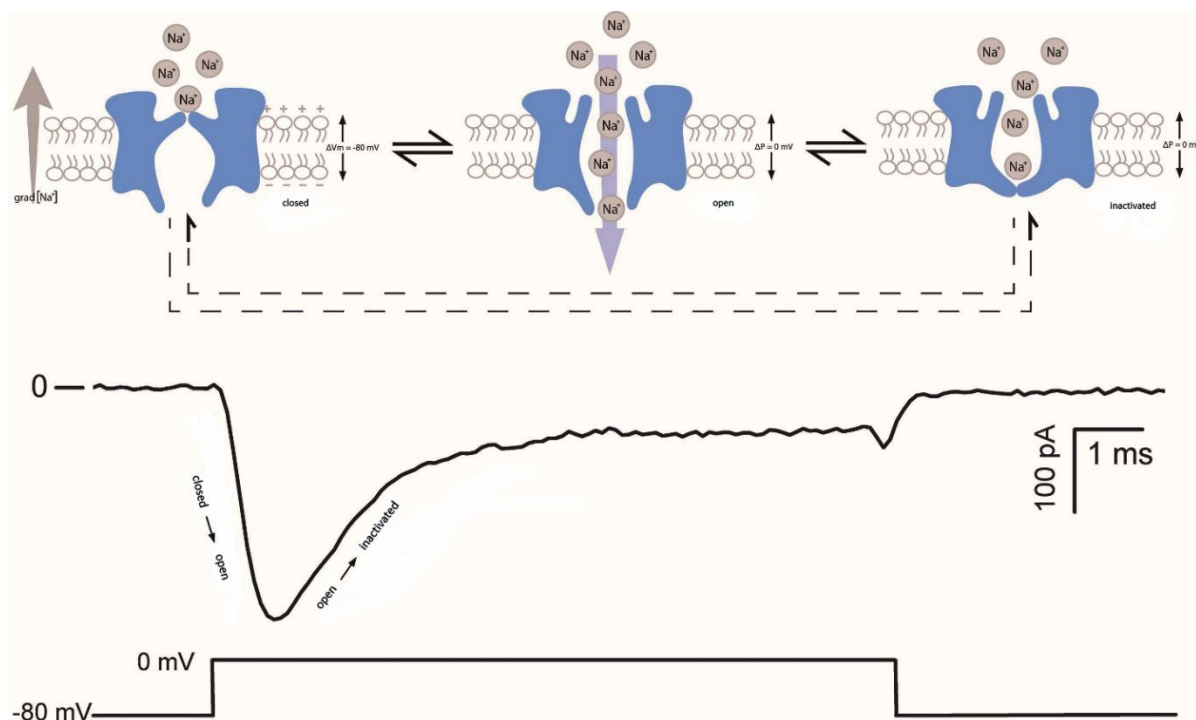


Figure 1.1. Transitions between ion channel states. The three main states are represented: closed, open and inactivated in a sodium (Na^+) voltage-dependent ion channel (image courtesy of Andrés S. Gandini and Carlos Manlio Díaz García).

Cnidarians venoms are also rich in toxins that act upon sodium (Na^+) and potassium (K^+) voltage-dependent ion channels [58]. In cnidarians it has been recognized at least four classes of toxins that act upon potassium channels, while other three classes are known for sodium (Na^+) voltage-dependent ion channels [22]. Among the sites of interaction between hydrophilic neurotoxins and sodium ion channels, are four sites significantly involved in the extracellular loop of such channels [29]. In fact, some toxins from cnidarians, the α -toxins from scorpions and spiders, share common interaction with specific sites (e.g. the known site 3), preventing the fast inactivation of the channel [59], thus decreasing the time of the inactivated state, hence provoking overexcitability of the channel [60-62] (Figure 1.2).

On the other hand, the β -toxins from scorpions and some arachnids' toxins bind to site 4 of the sodium channel, enhancing activation of the channel at lower the threshold in which normally the channel is in the closed state (Figure 1.2) [35, 59, 63]. It has been demonstrated that the increasing of the channel activity induced by toxins, extends the action potential period. The physiological consequences could be an increasing in the rate of firing that leads to more active transmission in the neuromuscular junction [64]. Finally, this effect could conduct to paralysis and even death of the inoculated organism [29]. The $\alpha\beta$ -toxins from scorpions constitute an example of two different strategies of neurotoxins that modulate ion channel, through the interaction with different sites, which affects the opening probability. Neurotoxins from scorpion show a great but venom-specific diversification, adopting the CS $\alpha\beta$ motif [50].

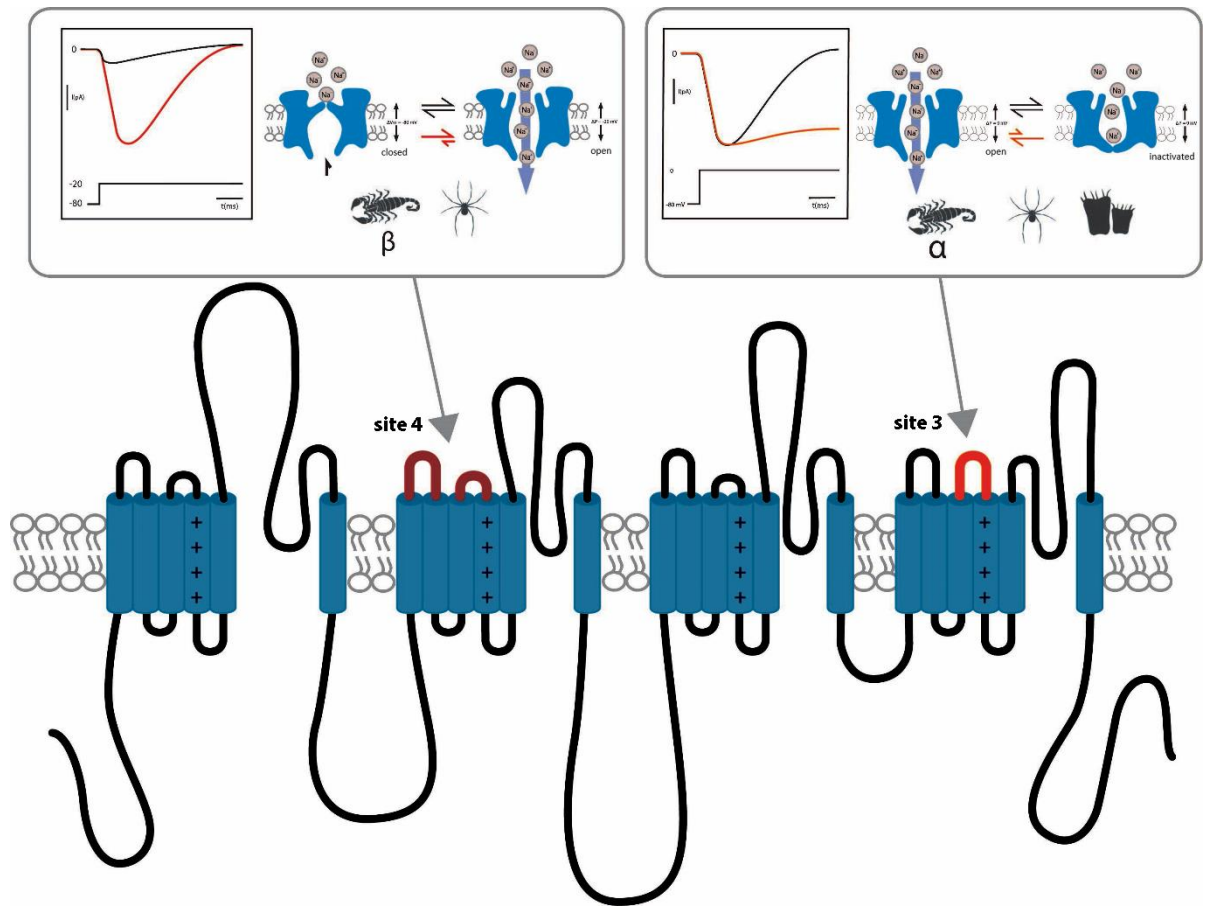


Figure 1.2. Toxins as modulator of site 3 and 4 of the sodium (Na^+) voltage-dependent ion channel. The effects of neurotoxins secreted by some organism are represented, which act on site 3 (scorpions α -toxins and some toxins from spiders and cnidarians) and site 4 (scorpions β -toxins and some toxins from arachnids) of the sodium (Na^+) voltage-dependent ion channel (image courtesy of Andrés S. Gandini and Carlos Manlio Díaz García).

Besides, other toxins can modulate other type of ion channels sensitive to pH, known as acid-sensing ion channels [65] and transient-receptor potential channels [66]. Toxins from snakes have been also found to act upon acid-sensing ion channels showing a high potential as analgesic [67]. Toxins from *Conus* are known for interacting with the pore-forming α -subunit of Na^+ , K^+ , and Ca^{2+} channels, which comprises three different *Conus* peptide families capable of modulating voltage-gated sodium channels [68]. Among these are included the μ -conotoxins that are channel blockers, the μO -conotoxins that inhibit Na^+ channel conductance, and the δ -conotoxins that delay or inhibit fast inactivation [68]. The μ -conotoxins and δ -conotoxins also modulate the channel binding to specific sites 1 and 6, respectively [69]. Besides, *Conus* produces other toxins known as ω -conotoxins that block the voltage-gated calcium channels [68, 70]. Moreover, the venomous repertoire of cone snails also includes the κ -conotoxin and the α -conotoxin as antagonists of the neuromuscular nicotinic acetylcholine receptors (nAChR), which causes paralysis of prey

like fishes [68]. Snakes also produce neurotoxins known as Cysteine-Rich Secretory Proteins (CRISP), which inhibit a variety of ion channels like Ca^{2+} and K^{+} channels [71-74], cyclic nucleotide-gated channels [75-77]; also ryanodine receptor channels [78].

Cytolytic toxins and hydrolytic enzymes

Commonly, venoms contain cytolytic toxins and enzymes with catalytic activity capable of form pores as non-selective cation channels in the membranes, with capacity to destroy the cells. Among them, it should be mentioned the cnidarians Cytolins, which based in their polypeptide primary structure can be classified in four families [22, 79]. They show different molecular weight ranging from 5-80 KDa, hemolytic activity and differ in their affinity for membranes containing specific phospholipids [22, 80]. One of the most known pore forming toxin is the α -Latrotoxin (α -LTX) from venom of spiders of the genus *Latrodectus* (widow spiders) [81]. This toxin acts presynaptically, forming a non-selective pore to cation; thus, provoking neurotransmitter exocytosis mediated by Ca^{2+} influx [9]. Spiders from the Genus *Loxosceles* produce a venom secretion rich in hydrolytic enzymes that produces necrotic wounds [82]. Among them, phospholipases, collagenases, hyaluronidases and metalloproteinases are involved in the ascribed toxicity [83-86], although the necrotic effects are mainly attributed to sphingomyelinases [87-89].

Phospholipases and hyaluronidases are widely distributed among animals highlighting cnidarians, mollusks, insects, arachnids and reptiles, but have also been found in scorpions, fishes, cephalopods, spiders, and ticks [2]. Phospholipase A2 (PLA_2), have been extensively recruited into venoms, comprising Group-IA, G-IIA, G-IIB, G-III, G-IX, and G-XII PLA_2 scaffolds [90]. Some PLA_2 s show particular features, like G-III which have been recruited independently into four different venomous lineages [2]. In addition, those from sea anemones, lack phylogenetically relationship with any other known PLA_2 types [91]. Among the toxic effects described for PLA_2 are included antiplatelet, myotoxic, and neurotoxic activities [4, 92]. The neurotoxic activity could be associated or not to the own catalytic activity [93], rather than its ability to bind targets [94, 95]. Unlike PLA_2 , hyaluronidases exhibit lower diversity and its function in venomous secretion is likely involved in the degradation of the tissues, thus increasing the permeability to allow a more efficient diffusion of other venom components.

Protease Inhibitors

There are also toxins acting as protease inhibitors. Some of them, like the Kaliclodines family from sea anemones can act both as ion channel modulators and proteolytic enzymes inhibitors [96]. These toxins are also known as Kunitz-type serine protease inhibitors, since they have a Kunitz domain, showing structural and functional

homology with serine protease like trypsin. The dual function mentioned as potassium (K^+) voltage-dependent ion channel blocker and protease inhibitor in the same protein constitutes an advantage, affecting important physiological process, but preventing an early degradation [14, 96]. Another Kunitz-type toxin with dual activity, acting both as trypsin inhibitors and K_v channel blockers, have been described in the mygalomorph spider venoms [97, 98].

Kunitz-type toxins represent a highlighted example of convergent toxin recruitment and convergent molecular evolution to produce the same derived activity [2]. Thus, it is not surprisingly that Kunitz-type toxins are commonly found among venomous animals [2]. Kunitz-types peptides have been isolated and characterized in sea anemones [14, 99-101], wasps [102], scorpions [103], spiders [98], *Conus* snails [104] and snakes [54, 105]. Besides, Kunitz-types peptides constitutes major components of venom in ticks and hematophagous insects, acting as inhibitor of the blood factor Xa [106, 107]. The anticoagulant effect of such toxins is very important in the feeding of these hematophagous animals. The Xa factor is the trypsin-like proteinase of coagulation that catalyzes the conversion of prothrombin to thrombin [108]. Thus, blocking the Xa factor can uncouple the coagulation cascade allowing them to effectively suck the blood fluids.

Toxins targeting physiometabolic processes

Toxins are mainly effective in targeting very irrigated organs to produce fast and acute effects after inoculation, since they can be degraded or neutralized [2]. In this sense, toxins produce powerful but short-term effects, being useful as research tools to study the underlying mechanism of physiometabolic process. Among such target organs, beta cells (β -cells) from endocrine pancreas constitute an example [109]. It is noteworthy the neurotoxin TsTxV from the scorpion *Tytilus serrulatus* that increase the insulin secretion glucose-stimulated mediated by sodium (Na^+) voltage-dependent ion channel [110]. Moreover, ω -conotoxins has resulted suitable to characterized the roles of calcium (Ca^{2+}) voltage-dependent ion channel in insulinome cell lines, which show differences among the expression of calcium channels subtypes [111, 112]. Besides, toxins of cnidarians, scorpions and spiders, have been employed to study and characterize some type of K^+ channels involved in insulin secretion [113].

In addition, a low molecular weight fraction from the crude extract of the zoanthid *Zoanthus sociatus* was assayed in isolated rats β -cells, blocking in a reversible way the influx of calcium (Ca^{2+}) and subsequently decreasing the insulin secretion glucose-stimulated [114]. Moreover, the effects of an insulin deficiency caused systemic intolerance to glucose after intraperitoneally administration [114]. On the contrary, two low molecular weight toxins isolated from the Portuguese man-of-war *Physalia physalis* increased the insulin secretion in the same model previously mentioned in isolated rats β -cells [115]. The

general mechanism involved is unlikely associated with voltage-dependent ion channel, despite it was detected intracellular variation in Ca^{2+} , which is involved in the secretion of insulin [115].

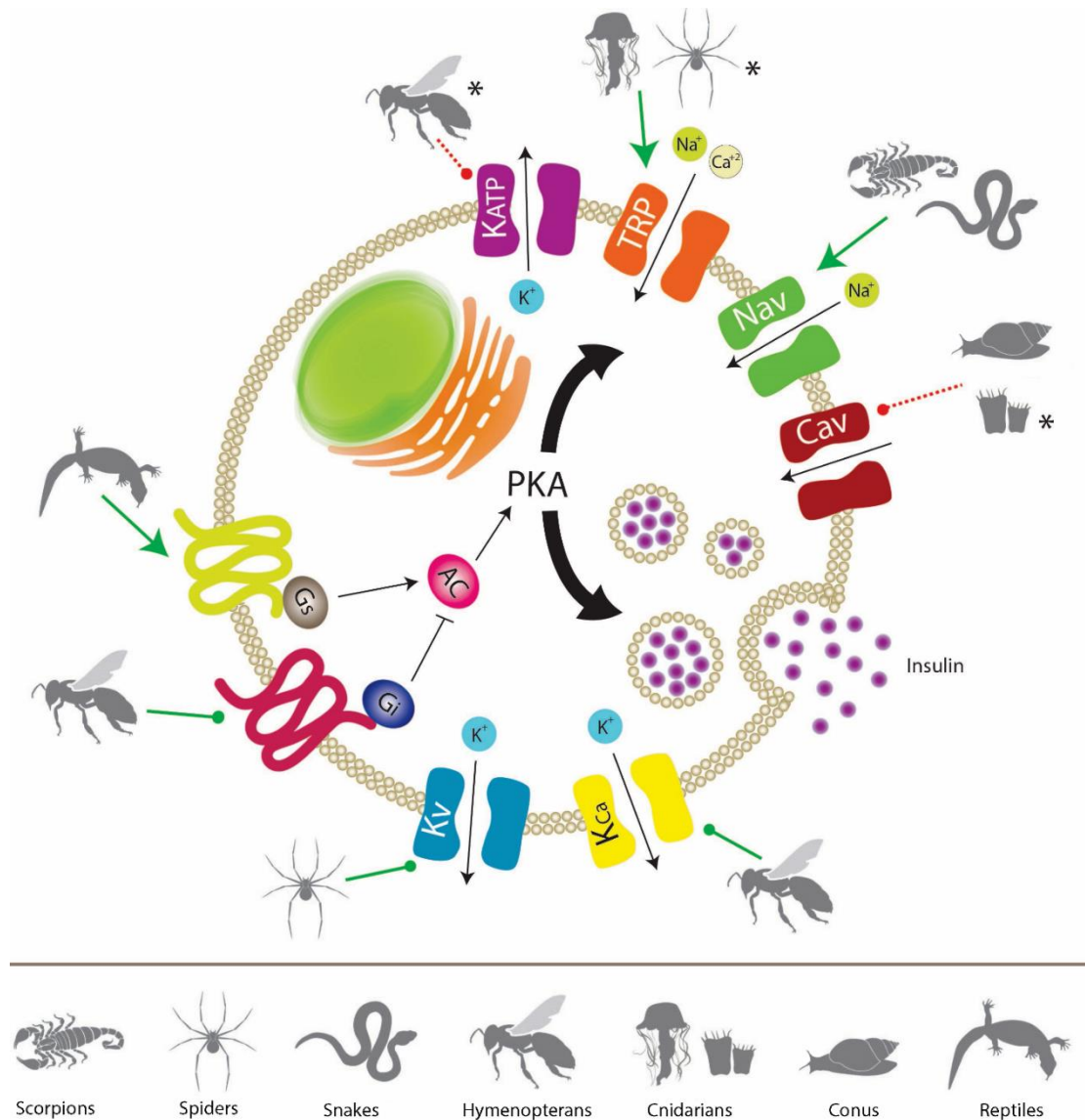


Figure 1.3. Animal toxins affect insulin secretion. Representation of the different effects of toxins from different animal lineages acting upon ion channels expressed in the pancreatic β -cells. The organisms with asterisk denote potential modulators of pancreatic β -cells physiology. The arrows indicate activation of the channel or receptor, whereas those lines ending in circle designate inhibition. The green color of the arrows highlights that the effect on the channel or receptor leads to an increase in insulin secretion, while the red color indicates its reduction. Image courtesy of Andr s S. Gandini and Carlos Manlio D az Garc a.

Insulintropic effect, is not restricted to cnidarians' toxins, since a toxin from the cobra *Naja kaouthia*, firstly described as cytolytic, increased the insulin secretion as a results of calcium cytosolic raise mediated by blocking of K^+ currents [116]. Besides, exendin-4, a

peptide isolated from the venom of the Gila monster *Heloderma suspectum* has more than 50% homology with the glucagon-like peptide (GLP-1). The last mentioned peptide GLP-1 belongs to the family of the Incretins, which are gut hormones secreted from enteroendocrine cells [117]. The mechanism of action involves interaction with G protein-coupled receptors (GPCRs) that increase cyclic AMP (cAMP), which potentiates insulin secretion and beta-cell survival [118].

In the Figure 1.3 it is shown an example of different animal toxins that can modulate insulin secretion through the interaction with ion channels expressed in pancreatic β -cells [109]. Among targets and enzymes involved in the depolarization phase it should be noted the ATP-sensitive K^+ channel (K_{ATP}), transient receptor potential (TRP) channel, voltage-gated Na^+ (Na_v), low voltage activated (LVA) Ca^{2+} channels (mainly L-type), allowing an increment of intracellular Ca^{2+} , thus insulin releasing by exocytosis [109]. Enzymes like protein kinase A (PKA) and Adenylate cyclase (AC) have also a relevant role in this pathway [119, 120]. Then, repolarization happen as a consequence of the outward current of K^+ through the delayed rectifiers (K_v) and Ca^{2+} -sensitive big conductance K^+ channels (BK) [109] (see Figure 1.3).

Immune response against animal toxins

In general, animal toxins are less immunogenic as their size decrease [121]. It is known that many of them are peptides, but sometimes toxins exhibit high molecular weight triggering severe systemic immunological response called anaphylaxis [122, 123], which can lead eventually to death. This phenomenon was previously described following the administration of sub-lethal repeated-doses of extracts and toxins from the cnidarians *Physalia physalis* and *Actinia sulcata* [122, 124], and the findings were awarded with the Nobel Prize in Medicine and Physiology in 1913. Besides, the human immune system can produce antibodies with cross-reactivity between cnidarian toxins, like some from *P. physalis* and *Chrysaora quinquecirrha* [125]. Furthermore, cnidarians toxins have been revealed as a source of target-specific immunomodulators. Indeed, a potassium channel toxin ShK from *Stichodactyla helianthus* lead to a promising drug-candidate ShK-186 to treat autoimmune diseases [126, 127] (see the section below of “Biomedical potential of toxins”).

The most immunogenic animal's toxins probably are those belonging to insects. Among them, hymenopterans venoms (bees, wasp) can commonly generate anaphylaxis in some hypersensitive individuals [128, 129]. These effects are associated with abnormal concentration of IgE antibodies in serum with high affinity to some venoms components. The symptoms can produce urticaria (hives), angioedema, shock and cardiorespiratory arrest. Toxic effects often arise once histamine is released by mast cells after IgE-mediated interaction [130, 131].

1.3 Biomedical potential of toxins

Toxins are of biomedical and pharmacological interest, since in a low concentration they mimic natural process with high specificity and affinity [10]. Toxins represent a source of useful tools to unravel complex physiological process, including nervous systems, the cardiovascular system, blood coagulation, homeostasis, the hormonal system, the complement system and also the immune system [132]. In fact, they have been considered most useful in an earlier stage of drug research in order to identify potential therapeutic targets, ion channels characterization, or explore physiological disorders [10]. However, toxins can be used directly as therapeutic agents [133], since venoms constitute already an approved source of drug-leads like analgesics, anti-hypertensives, immunomodulators, antiplatelet, pro- and anti-coagulant agents, fibrinolytic and also as anticancer candidates [134-137].

It is noteworthy, that some drugs or active pharmaceutical products derived from venom components are available on the market or are in process to be tested in pre-clinical or clinical trials [69, 134]. An example is the synthetic antihypertensive captopril derived from the venom of the Brazilian snake *Bothrops jararaca* that mimics the bradykinin-potentiating peptides effects [138]. Besides, strong analgesic called Prialt® (ziconotide), is also a synthetic venom-derived drug of ω -conotoxin MVIIA, a peptide from the venom of the marine snail *Conus magus* [70]. The mechanism of action of this drug is mediated via a blockade of N-type calcium channels [70]. In addition, the synthetic Exenatide (Byetta®) derived from exendin-4 from the saliva of Gila monster lizard was approved on the treatment of Type-2 diabetes [139]. Moreover, a chlorotoxin from the scorpion *Leiurus quinquestriatus* was initially used to characterized glioma-specific chloride currents, because its selectivity on glioma cell [140]. Then, its modified version TM-601 constitutes a promising alternative to the treatment of this diffuse form of brain cancer [140, 141].

On the other hand, ShK-186 [126] is derived from the toxin ShK from the sea anemone *Stichodactyla helianthus* [142] and represents one of the most competing venom-derived drug, specifically to control autoimmune diseases like rheumatoid arthritis, Type-1 diabetes mellitus and multiple sclerosis [143, 144]. The diseases happened as consequences of tissue destruction mediated by autoreactive (self-reactive) T lymphocytes [126]. Some diseases are associated with an overexpression of Kv1.3 channel, which is involved in the proliferation, migration, cytokine secretions of such T lymphocytes [126]. The synthetic version of this toxin ShK-186 acts selectively upon T lymphocytes potassium channel called Kv1.3 [126], modifying the immune response but with a minimum of adverse

effects [145]. Moreover, some subtypes of ion channel are also overexpressed in cancer, constituting a remarkable target to control cell proliferation mediated by toxins [146].

However, more efforts are likely addressing the find of new analgesic compounds at this moment. In fact, some venom-derived painkillers are in study resulting as promissory candidates [147]. Among them can be mentioned “the mambalgins”, two peptides isolated from the black mamba *Dendroaspis polylepis polylepis*, which act upon acid-sensing ion channels [67]. These peptides have shown a powerful analgesic effect, even more than morphine, but with less tolerance drug-induced [147]. Another analgesic peptide known as THA903 was isolated from the King Cobra *Ophiophagus Hannah* venom, also stronger than morphine showing rapid effect after oral administration in pre-clinical pain models [148]. Despite its potential in therapeutics, the diversity of toxins is still underexplored [69]. Another putative analgesic is being developed from *Conus catus*, a relative of the toxin from *C. magus* that resulted in the Prialt® product [147]. It is noteworthy that only the genera *Conus* comprise around 500 cone species (*Conus* spp.), accounting for more than 11 pharmacological classes [69], while spiders comprise around 14,000 described species, with an estimated 12 million of biologically active peptides [149]. To date, approximately 1500 reported toxins have been described from this megadiverse group of venomous animals [150, 151].

1.4 General methods in the discovering of new toxins

In general, the complete isolation and characterization of venoms comprise some step that can be grouped into: 1) sample collection and/or extract preparation, 2) purification or fractionation of venom components, 3) bioactivities assays and 4) structure elucidation. The bioactivity of venom or fraction, or pure toxins can be performed in parallel in any step of the process, which is known as bioassays-guided. The bioassays-guided are useful to know the bioactivity of the extract and the subsequently obtained fractions are tested in an *in vivo* or an *in vitro* model. Besides, bioassays-guided is useful to evaluate whether the protein or toxin of interest lost its activity during the purification process.

1.4.1 Isolation and characterization of cnidarians toxins

General protocols on cnidarian venom extraction

In the case of cnidarians, the discovery of new toxins starts with the preparation of crude extract or exudate from the sample collected that contains the venom secretion. Firstly, the extract can be me prepared from the tentacles or the whole body of some individuals, applying blender in distilled water, which releases the toxins. These procedures have been applied mostly in sea anemones, but can be considered as more primitive than

those used from isolated nematocytes [142, 152-155]. However, some experts are still using such methods due to the high yield of toxin. This makes sense because sea anemone has nematocytes in the gastrovascular cavity [156, 157], thus notable amount of toxins can be lost if the extract is only made from exudate or tentacles. However, for a cleaner starting solution, or better comprehension of toxins secretion, the mucus or exudate are preferred [45, 158]. In this method, the specimens are introduced in a recipient with distilled water and gently squeezed. Then the mucus is concentrated and recovered by freeze-drying, subsequently resuspended and centrifugated in repeated cycles to obtain a supernatant containing the toxins.

Unlike sea anemones toxins, the toxins from jellyfish are less stable, requesting the quicker preservation of tentacles, exudate or nematocytes to avoid degradation [155]. In general, some strategies can be used to stop the enzymatic machinery both in jellyfish and sea anemones. Specimens, exudate or part of the animal like tentacles can be kept in ice, or bring specimens to the lab into plastic wear containing sea water to make the procedures of venom extraction [155, 159]. Temperature control is one of the most important variable to consider, representing a suitable solution for transporting or preserving the venom [155]. Hence, can be used ice, dry ice, lyophilized material, starting material stored at -20°C and snap-freeze tentacles in liquid nitrogen and then stored them at -80°C [160-165]. However, sometimes nematocytes remain intact and some other procedures are required to release the toxins. Among them, freeze-shaken, pestle and mortar, autolysis and some alternatives to autolysis like the use of glass beads have been successfully applied [155]. The last one mentioned, has been referred as one of the most effective regarding yield, time-saving and avoidance of contamination [155]. Sonication of extracts or isolated nematocytes are subsequently included in the protocol for better yield, followed by some centrifugation steps. After the extraction procedures, the solution containing the venom or the lyophilized material can be kept at lower than -20°C until use [162, 163, 166, 167].

At this end, the supernatant is concentrated under vacuum or the lyophilized material resuspended to determine the protein concentration. Total protein content can be estimated by Bradford method [168], or by measuring maximum absorbance of the peptide bond and aromatic residues at 220/280 nm [169], respectively. In the last method, bovine serum albumin is used as a reference standard. Other methods to estimate protein concentration include the Nanodrop™ [170] and the available Pierce BCA assay kit commonly used [115, 162-164, 171-173]. Besides, older methods such as the Lowry [174], or other more simple based on ultraviolet spectra like the Wadell's method [175], are also useful. Finally, prior to purification step, it is usually performed one dimension SDS-PAGE electrophoresis, in which the components can be separated in gel according to its sizes, given an idea of the complexity and the molecular weight composition of the fraction [155].

General procedures in the purification of cnidarians toxins

The process is then followed by fractionation, and usually combined with purification steps for the activity test, known as bioassays-guided chromatographic purifications [158]. The most employed method in the fractionation of cnidarians toxins combined with gel filtration and reversed-phase chromatography (RP-HPLC) [176], mainly for the peptidomic analyses [158, 177]. The gel filtration constitutes a low pressure chromatographic technique, useful in the first step of purification [176], allowing the separation of the component per size. In this technique, the purification column is packed with different matrixes depending of the purpose. Among them, Sephacryl and Superose are used for wide ranges [178], Superdex for high resolution [114, 115] and Sephadex for a variety of matrix sizes [179, 180]. Usually, the extract previously dissolved and centrifuged is loaded onto Sephadex G-50 M. The system can be coupled with UV/conductivity detector to monitor elution at $\lambda=220/280$ nm in 0.1 M ammonium acetate or acetic acid at a low flow rate (1-10 mL/min), fraction collector and paper recorder [176, 180].

However, other systems have been used, specifically for jellyfish toxins. Molecular exclusion matrix as Sephadex G-100 column pre-equilibrated with Tris-HCl buffer [180] or Sephadex G-200 then conducted to a Sephacryl S-200 HP can be used. Superdex has also been successfully applied, given a wide range of cut-off size exclusion. Among the systems and matrixes used, it can be mentioned the Superdex 75 column pre-equilibrated with standard mixture (SERVA) [181]. In addition, the Superdex 75 column was also employed with ammonium acetate as pre-equilibrated and with elution buffer solution [115]. Besides, other studies used the Superdex 200 10/300 GL column in Fast Protein Liquid Chromatography (FPLC) [182]. This systems is suitable to obtain compounds between 600 KDa–10 KDa [182], where most of the cnidarians toxins belong [14, 176]. However, some known toxins from cnidarians, specifically jellyfish toxins, exhibit high molecular weight over 40 KDa, even 100 KDa [169, 183]. In addition, Superdex S-200 column also resulted in a system monitored at 280 nm using PBS as eluting solution [184]. These methods usually have been combined with pH gradient ion-exchange chromatography and reversed-phase chromatography, resulting both effective for sea anemones and jellyfish toxins [155, 176].

These purification steps are then followed by reverse-phase (RP) High Performance Liquid Chromatography (HPLC), which is generally employed for peptide toxins separation before mass spectrometry analyses [155, 176]. Unlike previous chromatographic steps, RP-HPLC provides higher resolution combining such reverse-phase with a High-Pressure Liquid Chromatography system [185-187]. The reverse phase is associated to a mobile phase significantly more polar than the stationary phase. In other words the stationary phase is hydrophobic, which is useful to separate low retained compounds [188, 189].

Then, the fractions obtained are loaded onto reversed-phase HPLC column containing C18 matrix, preequilibrated with ammonium acetate or acetic acid. Components are eluted using a linear gradient of acetonitrile:trifluoroacetic acid (ACN:TFA), monitored at 220/280 nm [114, 115, 176]. In parallel with the purification process, the bioactivity of each fraction can be tested and pooled. The use of a combination of gel filtration, ion-exchange with RP-HPLC approaches resulted as some significant improvements in the purification of venoms. Unlike other venomous animals like scorpion and spiders in which multidimensional chromatographic step have been so far applied [190-193], in sea anemones, commonly have been used only a combination of gel filtration and reversed-phase chromatography [158, 176, 177]. The use of multidimensional chromatographic step including filtration, ion-exchange and RP-HPLC approaches have been also suitable for the purification of cnidarians toxins, as well as for other venomous animals [176].

Proteomics in the study of venom

Proteomics aim the study of the proteome, which is associated with a wide range of proteins from a tissue and its function [194-196]. To reveal the proteome, some approaches like mass spectrometry analyses are required. There are some methodologies of mass spectrometry analyses to assist the characterization of the whole proteome or a specific fraction or protein. In general, mass spectrometer is needed, which ionizes the sample and then the resulting ions are sorted by different physical-chemical properties like mass-to-charge ratio (m/z) [197-199]. Then the obtained spectra based on mass-to charge ratio are analyzed with mathematic algorithms and further identification can be conducted by homology using statistic-based bioinformatic tools [200-203]. Finally, peptides and proteins can be reconstructed and identified against knowledge databases like Uniprotkb/swiss-prot [204, 205] or the compilation of databases from the Database resources of the National Center for Biotechnology Information (NCBI) [206], using MASCOT algorithm [200, 207], Protein Pilot protein identification software v4.5 (AB SCIEX) [208], PEAKS [203] or MaxQuant proteomics software package [201, 202].

Thus, mass spectrometry has become a mandatory method in both the identification of proteins and discovery of candidates as new drugs [209]. Probably, the most used spectrometer in the study of venoms are matrix-assisted laser desorption-ionization (MALDI) and electrospray (ESI) [115, 118, 155, 210, 211], resulting suitable for proteinaceous components, since the relative low ionization energy employed and the wide m/z range detection [212]. Peptide Mass Fingerprinting and protein identification by database searching of tandem mass spectrometry (also known as MS/MS or MS^2) are among the most used strategies [158, 176, 213]. The first method consists of a mass spectrum from a mixture of peptides digested usually with trypsin, which then is used as a fingerprint to identify proteins

[214]. In the second strategy more than one step of mass spectrometry selection is needed, generating some fragmentation level between the stages [215]. These strategies have shown some limitation for the study of animal venoms, with the exception of animals like snakes where enough amount of venom can be obtained [211]. However, both strategies have also been recurrently applied in the discovering of cnidarian toxins with success [158, 176, 216-219].

Because of the limited amount of venom available from some species, analyses at the nanogram-to-microgram scale have been implemented. Proteomic and mass spectrometry analyses are evolving toward high-throughput analyses, given insights in the large-scale of the proteome with a minimum amount of venom [69]. Specifically, nano liquid chromatography (nanoLC-MS/MS) have been already improved in the study of cnidarians toxins [155, 167]. Techniques that combine high performance liquid chromatography with mass spectrometry are also known as shotgun proteomics [220-222]. In this sense, LC-MS/MS is likely the most powerful techniques in the field of proteomics, allowing the identification of several proteins out of the complex protein mixtures with high resolution and confidence [155]. However, some experts consider the electron-transfer dissociation as the most promisor technique to unravel large venom peptides [69, 223].

In the identification of venom from jellyfish, other strategies that resulted effective included the LC eluting system coupled with Finnigan LTQ linear ion trap MS using electrospray ionization [166, 167], as well as NanoSpray II ionization source of a QSTAR Elite Hybrid MS/MS System [224]. It is noteworthy that both the number of protein identified and the confidence could be notably improved whether LC-MS/MS techniques are combined with transcriptomic approaches [166, 225]. Another study based on LC-MS/MS employed orbitrap as the spectrometer shed light on the toxins composition from nematocysts of the Atlantic jellyfish *Olindias sambaquiensis* [170]. In the last study mentioned, the system comprised an EASY-nLC II (Proxeon) nano LC system connected to a Thermo Scientific Orbitrap Velos Pro mass spectrometer. In addition, some of these methods can be also applied without previous protein separation known as whole venom fingerprinting, given a profile of the proteins contained in the whole venom [212]. The mass spectrometry analyses can be combined with one or two-dimensional gel electrophoresis allowing the separation of the compounds prior to the mass spectrometry analyses [155, 212]. In this sense, it can be also called gel-based or gel-free mass spectrometry analyses [226], both resulting suitable for venom characterization [155]. Nonetheless, regarding the amount of venom available, gel-free could be an advantageous option if the proper spectrometer is available.

Once the raw data profile is obtained (i.e. large data sets from high-throughput analyses), bioinformatic analyses are needed. Unlike MASCOT algorithm [200, 207], Protein Pilot protein identification software [208] or PEAKS [203], bioinformatic tools like MaxQuant,

Proteome Discoverer and X!tandem are recommended as database search algorithms [201, 227] for such proteomic data. Functional annotation is also needed to have a general overview about Biological Process, Cellular Components and Molecular Function of the proteome and the main enzymatic pathway involved. For this purpose, one of the most commonly bioinformatic tools used is the Blast2Go software [228]. Blast2Go generates the functional annotation of the proteins identified through the Gene Ontology (GO) (<http://www.geneontology.org/>) [229, 230].

1.4.2 Isolation and characterization of cnidarians toxins

Sequencing Platforms in NGS

High-throughput approaches have been formerly known as “next-generation sequencing (NGS)” tools comprising techniques that address large-scale genome and transcriptome sequencing, even though DNA-protein interactions known as ChIP-sequencing and epigenome characterization are also included [231]. NGS provides invaluable tools to explore new horizons in the knowledge of different levels of expression from genes to protein. Currently, NGS platforms allow the sequencing of the whole genome or total RNA (transcriptome) for affordable price, which is trending to fall even more. Among the available platforms for sequencing it can be cited Single-molecule real-time sequencing (Pacific Biosciences), Ion semiconductor (Ion Torrent sequencing), Pyrosequencing (454), Sequencing by synthesis (Illumina), Nanopore Sequencing and Chain termination (Sanger sequencing) [232, 233]. However, some authors do not consider as NGS those sequencing technologies based on the classical Sanger’s strategy of DNA sequencing [234], which corresponds mainly to 454-pyrosequencing (Roche), to paired-end (Illumina-Solexa) or to semiconductor (Ion-Torrent™) sequencing strategies [234].

To characterize each platform some variable such as read length, single-read accuracy, number of reads per run, time-consuming and price, must be considered as a ratio between advantages and disadvantages. All methods have a relative good accuracy and are also fast apart from the SOLiD sequencing. This method provides the worst read length, ranging from 35 to 55 base pair (bp), but it is cheaper and reliable for the sequencing of palindromic sequences [235]. On the contrary, Sanger sequencing is expensive and the preparation of the sample is a time-consuming process, but the sequencing read length is better (400 to 900 bp) [236]. Nonetheless, to overcome this limitation, some methods and sample preparation kits are available or in development [237].

On the other hand, real-time sequencing provides the longest read length (average 14,000 bp) [238-240], higher than 10,000 bp up to 40,000 bp. Besides, the cost is low and sequencing process is also fast, but the single-read accuracy with 87% [241] is one of the

worst, while throughput is moderate [241, 242]. Nanopore sequencing is also fast, the read length up to 500 kb not dependent on the device, but dependent on the library preparation. The cost is relatively higher and the throughput lower than in other platforms. Pyrosequencing shows good parameters except for the cost, but also high error rate, specifically for homopolymer errors [243]. Illumina sequencing is cheaper, with high accuracy and high sequence yield. The reads length ranging from 50 to 500 bp are acceptable, which varies depending on Illumina sequencer model. The cost per 1 million bases in Ion Torrent is relatively higher compared to Illumina, but the accuracy and the read length is similar (up 400 bp). The reads per run up to 80 million in Ion Torrent is higher or less than Illumina depending of the sequencer model, but like Pyrosequencing produces homopolymer errors [232, 233].

Transcriptomic approaches

NGS techniques combined with bioinformatic approaches constitutes the state-of-the-art tools to understand the structural and physiological basis of life through the sequencing in one run of several genes, transcripts and proteins. The higher level of knowledge revealed by these techniques will allow unraveling how some genes have evolved by producing new products, adopting new protein structures and consequently new functions [2]. In general, the findings resulting from NGS represent a speedway to shed light on the origin and evolution of life. Besides, the data generated constitute evidence to figure out similar genome background from closed phylogenetically species, like human and chimpanzee, which resulted in notably differences [244]. Moreover, in the same individual such as a butterfly, the appearance changes dramatically through the different stages into its ontogenetic development [104, 244-246]. These alterations are generated as consequences of some differences in the transcription of DNA to RNA during life stages. It should be noted, however, that unlike the transcriptome, the genotype is the same during the life of unique specimen. The whole RNA or transcripts resulting from an organism is known as transcriptome. The study of the transcriptome of an organism or a tissue provides the way to understand the basis of expression. Also, the transcriptome provides information on the relative expression of some genes and isoforms, giving insight on the role of its encoding proteins.

To generate a transcriptome usually the sample of interest is preserved in RNAlater RNA Stabilization Solution (Thermo Fisher Scientific, Waltham, MA). Then the total RNA is extracted and conducted to the sequencing process. Once the total RNA is sequenced, the assembly of millions of RNA sequences is needed. For the transcriptome assembly it is requested as many reads as possible, which then will be used for the transcripts reconstruction into overlapping contiguous strands, called contigs [244]. The most common

platforms for RNA-seq are Illumina and Roche 454, but Illumina is currently preferred because it could be cheaper and yields more sequences with increased accuracy. Briefly, in the assembly process [247], firstly raw reads are clipped to remove sequencing adapters. This procedure can be performed with TrimmomaticPE (version 0.35) and identical reads are then collapsed using fastx-collapser (FastX Toolkit version 0.0.14). Afterwards, collapsed sequences are used for the *de novo* transcriptome assembly with Trinity [248] and Oases [249] assemblers. To improve some metrics as contigs length, assembled sequences from Trinity and Oases were reassembled with CAP3, which is known as an Overlapping Layout Consensus (OLC) assemblers [250]. However, many reads can be found as overexpressed, so-called housekeeping genes, related to metabolism and other basic life-sustaining processes. Multiple alignments are needed once the transcriptome is assembled and before mapping and estimate gene expression levels. For this purpose, the open-access RNA-seq alignment software as Bowtie can be used. Then, Basic Local Alignment Search Tool (BLAST) is performed and the transcripts annotation revealed [244].

Transcriptome advances in the study of venom snake glands

Transcriptome analyses have significantly increased the number of snakes' toxins, both front fanged and colubrid [105, 251-256]. Transcriptomic approaches combined with proteomic analyses have revealed several compounds involved in the venom gland secretion. In the transcriptome, the amount of product is associated with genes expression, which in turn is related with the number of copies of a transcript. Some of these components can be more or less expressed and can be classified as major or minor components, depending of the concentration in the venom. However, each component can play specific or diverse role in the venom effect. Thus, any component must not be underestimated, since commonly some of them can facilitate the toxic effect of other, acting synergically.

Among major and minor components have been found the Phospholipase A₂, Snake venom metalloproteinase (disintegrin/metalloproteinase types II and III/reprolysin-type), Snake venom serine proteinase, Hyaluronidase, Phosphodiesterase, Cysteine-rich with EGF-like domain, Cysteine-rich secretory protein, Glutaminyl-peptide cyclotransferase, Protein disulfide isomerase, L-amino oxidase, Kunitz-type protease inhibitor, Acetylcholinesterase, Nucleotidase, Bradykinin potentiating, Myotoxin/crotamine; Lipocalin/lipocalin-like toxins, Vascular endothelial growth factor, Endothelin, Nerve growth factor, Vespryn (ohanin-like)/ohanin, Venom factor, Neurotrophic factor, Betanerve growth factor, Factor V, Factor X, C-type natriuretic peptides, ADAM, AVIT, complement C3, crotasin/beta defensin, cystatin, Kallikrein, LYNX/SLUR, C-type lectin, Galectin/galectin-like, SPla/Ryanodine, and whey acidic protein/secretory leukoproteinase inhibitor and Waprin [2, 4, 252, 256, 257].

1.5 Animals models in the characterization of toxins

Toxicity is an inherent feature of any substances associated with its capability to produce some damage in a tissue, cell or disengage physiological process, especially when causing death or serious debilitation [258]. The toxicity of a substance is related with their chemical-physics properties, hence with the process involve in the absorption, distribution and excretion; also with the dynamic interactions to its target. The toxicity of a substance is inversely proportional to the dose applied. Moreover, the effects induced also depend on the physiological state of the animal, the interaction with other physical or chemical agents and with the biotransformation process. Altogether, all these elements can regulate the excess or defect of the active principle; thus, modifying the toxic effects [259, 260].

Overview of toxicological tests

The capacity of a substance to produces damage in short-term, after one administration, or some administration in 24 hours is known to acute toxicity. To determine the acute toxicity, the quantification of the mortality percentage is needed, in a range from innocuous to lethal dose. The lethal doses (LD) are those that lead to death in the tested animals. Among the lethal dose values, it can be mentioned the LD₁₀₀, which the subscript means that this dose reaches the 100% of lethality in the population tested. Thus, are very useful the LD₂₅, LD₅₀, LD₇₅. The LD₅₀, known as median lethal dose, is a measure of amount of substance that kills half (50%) of the population tested. This parameter is commonly used to estimate the toxicity of a substance, since it is used as reference to compare with the known LD₅₀ values [260].

On the other hand, other techniques have been developed to study the toxicity of sub-lethal doses of a substance in the long-term. The study of toxic effects that can be maintained at a time, which corresponds to the tenth of the life of organisms, is known as chronic toxicity [261, 262]; while sub-chronic studies or repeated-doses studies are less time-consuming chronic tests [263, 264]. In general, these methodologies address the studies of the impact of toxicity in carcinogenesis, reproduction, parents' fertility and ontogenetic development.

However, testing toxic compound in animal requires some ethics protocols regarding its manipulation. The rules have been established since 1982, by the *Organization for Economic Co-operation and Development* (OECD) [265]. The guidelines for reduction of injuries and the number of animals used in different *in vivo* toxicological assays have been proposed as a general ethic rule among protocols. In this sense, there is guidance for better ethical use of animals in toxicology. These three "R" principles are Replacement, Reduction

and Refinement [266]. The aim of these principles promotes alternative models or use of methods that enable researchers to reach the aims using fewer animals and to minimize potential pain, suffering or distress. Alternative methods like cytotoxicity assays in cell lines are currently recognized as useful, but an assessment in animals is still mandatory before conducting tests in humans [267-269].

Behavioral assessment in pharmacological studies

Animal models are commonly employed in pharmacological tests in many laboratories. Through specific *in vivo* tests, it can be revealed the pharmacological profile of an unknown compound using pre-established patterns of drugs-induced effects as reference. These methods attempt the discovery for new drugs, either through the simple observation and interpretation of behavior. Each laboratory modifies the methods per their needs and resources. Usually, the methods are modified using cameras, photoactive cells, statistical packages and professional data analyses programs, with the aim of quantifying certain variables that may be associated with pharmacological effects. One of the most well-known and used systematic methods is that described by Irwin in 1964 [270, 271], commonly known as “Irwin test” in the field of pharmacology. These methods are especially suitable in the first step of new drugs characterization, which allow the discrimination among different pharmacological-agents classes. Drugs effects upon nervous and cardiovascular system could be easily inferred during toxicological tests, through observational assessment using conventional animals like rats or mice.

The intermittent observation method has also been used for the study of stimulant and sedatives drugs, especially in those with anxiolytic effect [272]. The sedative properties of substances are tested in rats or mice and the spontaneous motor activity is measured [273]. This term does not refer only to locomotion, it also includes stereotyped behaviors such as grooming, exploration (sniffing, scraping, two-legged lifting) and can be affected by factors such as light, temperature, animal strain [274, 275] and age [276]. Another method "Open Field" [277] is regularly used, whether automated or not. Open field method aims to quantify the behavior of the animal in an open space and measure variables such as motor activity, speed, locomotion [278-281]. Variation in exploratory behavior is considered a measure of the degree of curiosity of the mouse, and is suppressed in relatively low doses by some drugs like benzodiazepines [282].

At present, there is a great interest in neurodegenerative diseases. Among the most novel approaches to study these pathologies are the design of genetically modified animal models (knock-out mice). Knock-out mice are used in tests such as the elevated plus-maze and the mentioned open field, allowing to evaluate learning, memory, fear and the anxiety [281, 283-285]. In other animal models, drugs such as scopolamine are used to induced a

state of damage in the central nervous system [286, 287], while in others, ischemia is simulated by temporary occlusion of the carotid arteries [288]. Afterwards, these individuals are conducted to Morris water maze or passive avoidance test [289-291].

1.6 General aims

Cnidarians

Approximately 13,000 living cnidarian species have been described by taxonomists [292]. Despite its diversity, that constitute a great potential as a source for research or as models for therapeutic treatment [69], as cnidarians toxins still remain mostly unexplored [22, 155, 176, 293, 294]. Cnidarians are in fact the largest phylum of generally toxic animals, yet their toxins and venoms have not received as much scientific attention until 1990, comparatively to those of many terrestrial and even marine animals [295]. Sea anemones have a great potential as a source of peptide/protein toxins within cnidarians, in part because their toxins are considerably stable compared to other cnidarian toxins (e.g. jellyfishes). Sea anemones toxins include voltage-gated Na⁺ and K⁺ channels toxins, acid-sensing ion channel toxins, Cytolysins, toxins with Kunitz-type protease inhibitors activity and toxins with Phospholipase A2 activity. However, only less than 100 species of sea anemone have been examined for peptide/protein toxins, although more than 1000 species of sea anemones are recorded in the World [157]. Thus, this group represents an unexplored potential source of therapeutic compounds.

Cuba is the largest of the Caribbean island and the presence of large cnidarians colonies around the island is recorded up to 30 million of years ago [296], currently comprehending more than 270 species of cnidarians. However, less than 20 species have been tested for the presence of protein toxins or activity from the whole-body extracts. In addition, Portugal coastal areas represent a notably resource of cnidarians toxins but just a few studies have been performed to date [22, 155, 293], mainly in sea anemones [294], where several toxins have been characterized [22]. Even when cnidarians species from Cuba and Portugal constitute a great source of bioactive compounds, its toxicological properties and toxins diversity are still poorly known.

Therefore, one major goal of this project is to perform the proteomic characterization of proteins/toxins from Cuban and Portuguese cnidarians. Specifically, one aim is to perform proteomics analysis, combined with bioassays to explore the toxins/protein composition of semi-pure fractions from the zoanthid *Zoanthus sociatus*. In addition, this work aims to provide the first proteomic map from the unexplored sea anemone *Bunodactis verrucosa* combining shotgun proteomics and gel-based methodologies. Each methodology applied may also be evaluated for comparative purposes. Whenever possible, the properties and

mechanisms of action of the new toxins discovered will be discussed. Moreover, this study could allow the identification of new compounds with pharmacological interest from Cuban and Portugal cnidarians.

Snakes

In addition, terrestrial vertebrates like snakes have been so far studied mainly for biomedical interest. Snakes represent one of the most explored venomous animals as source of toxins. Proteomic analyses shed light about venom composition, but the next generation sequencing has increased exponentially the number of snakes' toxins [248, 251-253, 256, 257, 297-299], even the structure of the genes responsible of toxins-encoding genes. Altogether, integrated omics approaches provide a wide view about toxins diversity and the evolution process involved that lead to new products; thus, adopting new features as differences in the sequences, structure, folding and protein functions. Although, the number of protein/toxins production can be inferred from the relative level of expression of the transcripts, affecting tissue function and theirs toxicological and pharmacological properties. Distribution, dietary shift or selection predator prey may be related with such phenomena.

Cuba is the greater of the Caribbean Antilles and due to the island isolation for millions of years it represents an important biodiversity hotspot that arise from the isolation of the species influencing their evolution and speciation, specifically for snakes. The population of Cuban snakes is represented by 40 species, of which 37 species are native, grouped in five families [300, 301]. Unlike front-fanged snakes, Cuban snakes' secretion have not been considered of biomedical relevancy. However, there are some unpublished reports about clinical complication after the Cuban racer bites *Cubophis cantherigerus*. It is noteworthy, that this species has Duvernoy's gland rather than the known venom gland responsible for the lethal reputation of front-fanged snakes. Despite some authors maintain the distinction between the Duvernoy's gland and venom glands [302, 303], others consider indistinctly as venom gland both the mentioned Duvernoy's gland and venom gland from front-fanged snakes [304, 305]. Duvernoy's gland is homologous and the ancestral version of the venom gland of front-fanged snakes [304, 305]. Indeed, most conservative researches have recognized that some Duvernoy's gland secretions are toxic, provoking pain, swelling, and other effects after subcutaneously inoculation [302].

However, it is well known that colubrid venom contains most toxins families found in front-fanged snakes [254, 256, 306]. Even, other cephalic glands from reptiles not associated with the upper jaw can produce toxins draining their secretions into the oral cavity [307, 308]. A cephalic gland, known as Harderian gland has a connection to the mouth by way of the lachrymal duct, which carries fluid from such postorbital gland to the vomeronasal organ [309, 310]. However, this gland remain as "the last remaining large organ of

widespread distribution among vertebrates, which function still to be cleared” [311]. The likely role of the Harderian gland are: a source of saliva, pheromones, thermoregulatory lipids and growth factors; part of a retinal-pineal axis, photoprotective organ, a site of immune response and osmoregulation [311, 312].

To unravel the proposed functions of the Harderian gland, transcriptomic analyses will be performed in three species of native snakes from Cuba. General characterization of gene expression of the Harderian glands, toxins-encoding genes expression will also be revealed. The whole profile of the Harderian gland from the three Cuban colubrids constitutes the first high throughput analysis of such glands in snakes, even vertebrates. The study of Cuban snakes will allow us to increase the knowledge on the venom production, toxins-encoding genes and its functionality. Common/distinct patterns of proteins/toxins expression among such Cuban species will provides clues about the mechanism of venom expression and toxins distribution among species, which will allow determining the relative expression and its likely ecological roles for each species. Moreover, all the transcripts obtained can be used as references for future proteomic analyses. Since the study of its toxin-encoding genes, genetics and evolution are mostly unknown, the transcriptome draft will be used in future works to assess phylogenetic relationships among snakes and its genes/proteins, as well as for evolutionary purposes to assess the reconstruction of specific genes involved in toxins production and if they have evolved under positive selection.

1.7 Thesis outline

The theme of this research is to explore the toxins repertoire of two main groups of venomous animals. Specifically, the proteomics analyses will be addresses for two species of cnidarians belonging to subclass Hexacorallia, Class Anthozoa from Cuba and Portugal, while transcriptomic analyses will be assessed on the cephalic gland from three Cuban snakes.

In **CHAPTER 1** it was provided a general introduction of the thesis, leading to the general and the specific aims of this work. Moreover, an overview of the methodologies applied was included to facilitate the comprehension of the main results obtained in the subsequent chapters.

In **CHAPTER 2** the toxicological properties of a low molecular weight fraction from the *Zoanthus sociatus* was evaluated. Some purification steps were done and toxicological tests in mice. The composition of the fraction was partially revealed by spectrometry analyses.

In **CHAPTER 3** purification and mass spectrometry analyses were carried out to determine the composition of the low molecular weight fraction from *Z. sociatus*, previously

found as lethal to mice. Unexpected, microcystins-LR was identified by the first time in such species, and the second time in cnidarians.

In **CHAPTER 4** another methodology was applied to characterize for the first time a proteome map of the unexplored *Bunodactis verrucosa* from Portugal. For this purpose, shotgun proteomic analyses were done resulting faithful for the identification of several proteins and peptides into the whole-body extract. In addition, previous results were compared with another methodology applied based on two-dimensional gel electrophoresis combined with spectrometry analyses with MALDI-TOF/TOF.

In **CHAPTER 5** the Harderian glands from three species of colubrid snakes from Cuba belonging to different ecological niches were analyzed using whole transcriptome sequencing. The three colubrid snakes selected were: *Cubophis cantherigerus* (Cc), *Caraiba andreae* (Ca) and *Tretanorhinus variabilis* (Tv). The sequencing based on Illumina-HiSeq provided the first insight in the gland gene expression, which have been previously considered as the most enigmatic organ of vertebrates without a clear role. The most expressed transcript shed light of the Harderian gland function, but opened a new question because snake's toxins were detected in a significant level.

CHAPTER 6 delivers a general discussion of the whole study main findings and reveals future perspectives.

CHAPTER 7 provides the general conclusions of the thesis.

Insights into the Toxicological Properties of a Low Molecular Weight Fraction from *Zoanthus sociatus* (Cnidaria)

Dominguez-Perez D, Diaz-Garcia CM, Garcia-Delgado N, Sierra-Gomez Y, Castaneda O, Antunes A: **Insights into the toxicological properties of a low molecular weight fraction from *Zoanthus sociatus* (cnidaria).** *Mar Drugs* 2013, **11**(8):2873-2881

2.1 Abstract

The phylum Cnidaria is an ancient group of venomous animals, specialized in the production and delivery of toxins. Many species belonging to the class Anthozoa have been studied and their venoms often contain a group of peptides, less than 10 kDa, that act upon ion channels. These peptides and their targets interact with high affinity producing neurotoxic and cardiotoxic effects, and even death, depending on the dose and the administration pathway. Zoanthiniaria is an order of the Subclass Hexacorallia, class Anthozoa, and unlike sea anemone (order Actiniaria), neither its diversity of toxins nor the *in vivo* effects of the venoms has been exhaustively explored. In this study, we assessed some toxicological tests on mice with a low molecular weight fraction obtained by gel filtration in Sephadex G-50 from *Zoanthus sociatus* crude extract. The gel filtration chromatogram at 280 nm revealed two major peaks, the highest absorbance corresponding to the low molecular weight fraction. The toxicological effects seem to be mostly autonomic and cardiotoxic, causing death in a dose dependent manner with a LD₅₀ of 792 µg/kg. Moreover, at a dose of 600 µg/kg the active fraction accelerated the KCl-induced lethality in mice.

2.2 Introduction

The phylum Cnidaria is an ancient group of predominantly marine simple animals that comprise over 11,000 extant species [292], which share a common diagnostic feature: the cnida [313]. The cnida is a subcellular organelle-like capsular with eversible tubules [314, 315] that contains the cnidocysts cells (also called cnidae). Of the three categories of cnidae (nematocysts, ptychocysts, and spirocysts), only nematocysts are found in all cnidarians [313]. This type of cnida may be associated with the production, discharge and inoculation of venoms in all cnidarians [156], whereby cnidarians are considered as the largest phylum of generally toxic animals [295].

Currently around 250 compounds from cnidarians have been identified including peptides, proteins, enzymes, protease inhibitors and non-proteinaceous substances [22]. Most cnidarians toxins have been successfully isolated from class Anthozoa, particularly from sea anemones, which is partly due by the stability of their toxins compared to jellyfish toxins [316]. To date, at least 191 proteins from sea anemones are recognized without ambiguities, considering the complete peptide sequences (or the information obtained by the translation of coding sequence submitted “CDSs” to GenBank database) and more than 80% deduced amino acid sequences for proteins over 10 kDa [317].

Most of these toxins correspond to peptides that act on voltage-gated sodium (Na_v) and potassium channels (K_v), whose molecular weights are between 3.5–6.5 kDa and 3–5 kDa, respectively [14]. These toxins also seem to have an universal distribution within the group, since all species tested have been found to contain toxins that are lethal or paralytic to crabs [14]. This claim is well supported by the number of toxins (62 for Na_v and 28 for K_v) characterized [22]. Moreover, the new Na_v toxins from *Aiptasia diaphana* [318] and the two novel type 1 sea anemone K_v toxins from *Bunodosoma caissarum* [319] should be added to this list.

Despite sea anemones being the best studied in the phylum, the order Ceriantharia, Corallimorpharia and Zoanthiniaria, which are closely related to the order Actiniaria, have stayed unexplored regarding the presence of low molecular weight toxins. There are only few reports on *Zoanthus sociatus*, an organism belonging to the order Zoanthiniaria, which is known to present organic compounds such as palytoxin [320] and some alkaloids affecting platelet aggregation [321]. Other biological properties have been described for *Z. sociatus* preparations, for example, an antifilarial activity by a chloroform-methanol extract [322] and an inhibitory effect on Ca^{2+} influx in rat β -cells by a low molecular weight fraction [114]. The latter suggest that there are unraveled biological activities in *Z. sociatus* that could account for its toxicity *in vivo*. In the present study, we obtained a low molecular weight fraction from

Z. sociatus and assessed its toxicological properties in OF-1 mice. A dose-mortality curve was constructed and conspicuous toxic symptoms were monitored and discussed.

2.3 Results and Discussion

Z. sociatus crude extract was subjected to a Sephadex G50 gel chromatography and the elution of its components was monitored by absorbance at 280 nm. The chromatogram exhibited various peaks that were pooled in four major fractions (Figure 2.1). The fraction ZsG50-III contained the most prominent peak and was selected for further studies because it has been demonstrated to contain low molecular weight toxins acting on ion channels in other studies on sea anemones using a similar methodology [323].

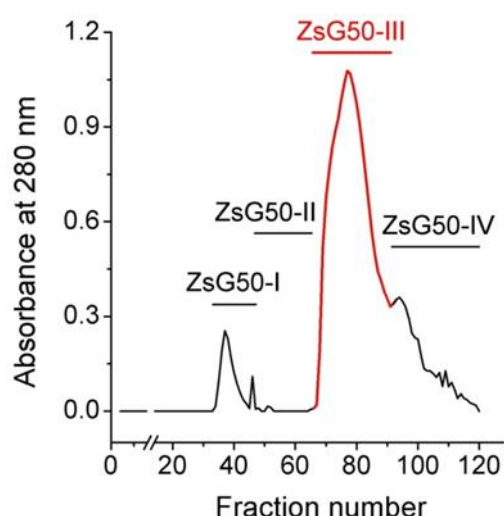


Figure 2.1. Sephadex G50 gel filtration chromatogram of *Z. sociatus* crude extract. Fractions in the chromatogram at 280 nm, obtained by Sephadex G50 gel filtration, were pooled in four major fractions. The fraction comprising the most prominent peak was named ZsG50-III and used for further toxicological studies.

The presence of signals detected by matrix assisted laser desorption/ionization time-of-flight/time-of-flight MALDI-TOF/TOF mass spectra (MS) corresponding to m/z ratio from 700 to 6000 Da. The highest relative abundance corresponds to compounds below 1000 Da (Supplemental Figure S1) that do not seem to be peptides (Supplemental Figures S2 and S3). However, some minor peaks between 2000 and 4000 Da captured and analyzed by MS/MS analysis in reflector positive mode showed typical fragmentation of peptides (Supplemental Figure S4). Molecular weights of these peptides are in the range reported for various toxins; however, blast analysis showed no significant similarity with any toxins from the UniProt database. Further procedures in the isolation of pure peaks is required in order to eliminate possible interferences in the detection of peptide signals by the main metabolites in the fraction.

To estimate the range of toxicity of the ZsG50-III fraction was started at a dose (150 µg/kg) and monitored the symptoms increasing in three-fold steps, until lethality was observed. Mice inoculated intraperitoneally with 150 µg/kg of the fraction of interest showed the same typical grooming activity of controls after injection. In the 450 µg/kg dose, animals decrease their exploratory activity 10 min after injection and remained near the walls of the cage. This symptom was accompanied with disordered breathing. At 1350 µg/kg these toxic effects were observed in less than a minute after inoculation, causing spasms, palpitations, convulsions and dead after 2 min. These effects were observed before sudden dead, which was preceded by dyspnea and reduced motile activity, suggesting cardiac arrest as the main cause of lethality. Certain drugs have been reported to cause respiratory and cardiovascular complications before cardiac arrest in mammals [324], including some cnidarian toxins [100]. It may be possible that the lethal effects were caused mainly by non-peptide toxins abundant in the fraction with molecular weight below 1000 Da. However, the presence of peptide toxins that can be acting synergically should not be discarded.

The acute toxicity results of five doses selected in the range between the two higher doses assayed in the preliminary test is shown as the percentage of lethality *versus* the dose plot fitted to a dose-response sigmoid curve (Figure 2.2A), which LD₅₀ was 792 µg/kg and the slope factor was 16.6 (expressing the dose as mg/kg). Toxicological effects appeared in less than 5 min after inoculation in mice from all groups; recovery, however, delayed proportionally to dose in those groups where lethality was not absolute and the time to death decreased in an exponential fashion (Figure 2.2B). It is worth to mention that 33 and 50% of mice inoculated with the two higher doses, presented fecal and urinary incontinency, perhaps because of relaxation of sphincter smooth muscle.

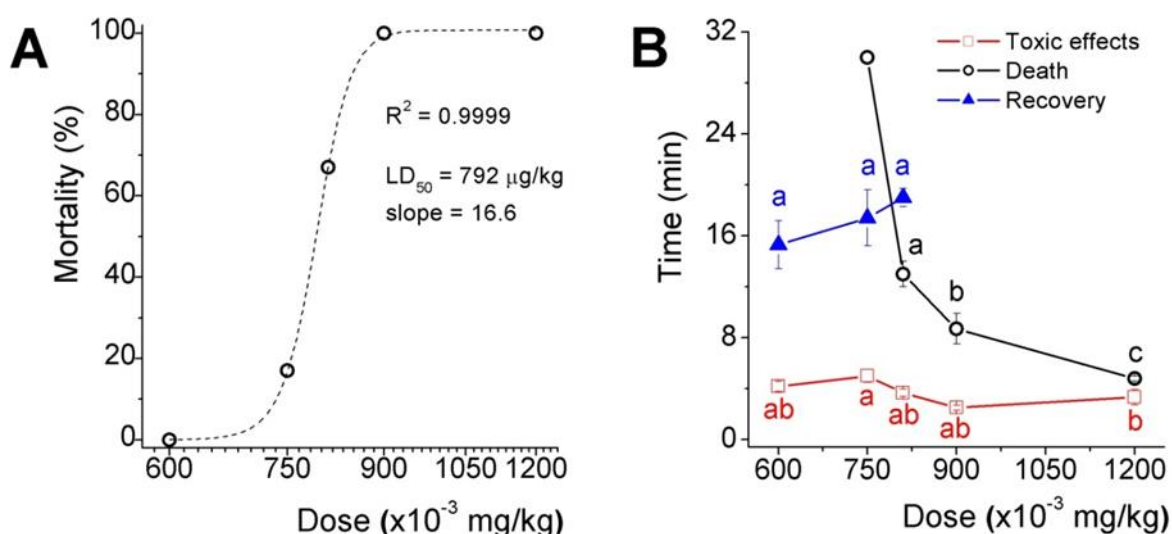


Figure 2.2. Acute toxicity assay of the low molecular weight fraction (ZsG50-III) from *Z. sociatus* crude extract.(A) Mortality *versus* dose curve. The plot shows a sigmoid equation fitting with the LD₅₀ and the slope factor for the low molecular weight fraction from *Z. sociatus*; (B) The plot shows the time to listed events at each dose. The appearance of cardiovascular symptoms was common to all doses, and around 5 min after inoculation. At the lower dose, there was a time window of 10 min approximately between the onset of these effects and their disappearance. Symbols represent mean $\pm$ SEM. Letters represent statistical differences for a $p < 0.05$.

Considering that an impairment of Ca²⁺ fluxes has been previously reported for low molecular weight compounds in *Z. sociatus* crude extract in excitable cells [114] and that voltage dependent Ca²⁺ channels are relevant to cardiac function [325], we decided to explore the effect of ZsG50-III fraction on KCl-induced cardiac arrest in mice, as it is known that hyperkalemia can produce cardiac arrest [326].

Seven mice were inoculated with KCl (1000 mg/kg) and six of them (86%) presented sudden death after 9.2 min in average (Figure 2.3). A similar dose of NaCl was applied to another group of six mice and no lethality was observed suggesting that cardiac arrest was mediated by high K⁺ instead an hyperosmotic load. To analyze if the fraction ZsG50-III could modify the lethal effect of KCl, we used a non-lethal dose of the fraction (600 μ g/kg) and compared the times to cardiac arrest respect to the KCl only group. All mice treated with KCl and ZsG50-III died in an average time of 6.1 min, which indicated that the fraction of interest accelerated KCl-cardiac arrest. Moreover, the controls with NaCl in this condition ($N = 6$) did not presented mortality and the symptoms were similar to the observed in the acute toxicity test for the same dose of the fraction in physiological saline.

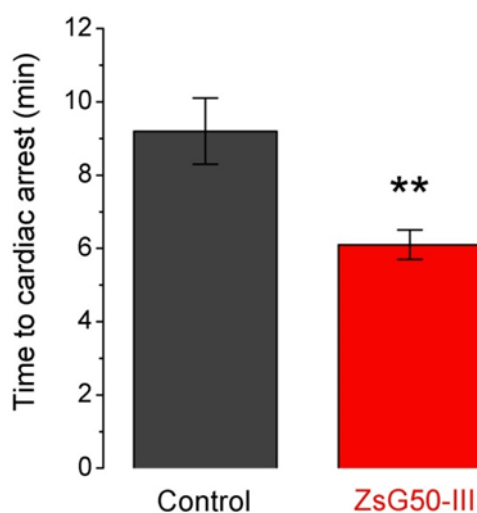


Figure 2.3. The low molecular weight fraction ZsG50-III accelerated the KCl-induced time to cardiac arrest. Bars represent the time to cardiac arrest after inoculation of a lethal dose of KCl (1000 mg/kg) in controls ($N = 6$) and simultaneous administration of KCl and 600 μ g/kg of the low molecular weight fraction in the ZsG50-III treated group ($N = 7$). Bars represent mean $\pm$ SEM. ** $p < 0.01$.

The acceleration of KCl-induced cardiac arrest by ZsG50-III suggests that the fraction could enhance the KCl-mediated cardiac dysfunction. It has been reported that *Z. sociatus* crude extract contains low molecular weight compounds that inhibit Ca^{2+} influx to pancreatic beta cells and impairs glucose tolerance in rats [114]. Interestingly it is well known that overdose of Ca^{2+} channel blockers cause cardiovascular failure [327, 328] but also hyperglycemia as a result, in part, of insulin secretion impairment [329]. In the light of previous studies and our results, a malfunction of Ca^{2+} fluxes could account for the observed cardiac toxicity during administration of high doses of ZsG50-III. However, further efforts should be made to unravel the exact mechanisms and to identify the responsible molecules.

2.4 Experimental Section

General Procedures

Z. sociatus were provided by the National Aquarium of Cuba, La Habana, Cuba. The zooids were brought to the Laboratory alive and kept in clean seawater until crude extract preparation. Sample collection and crude extract preparation. Briefly, specimens were cut in small pieces after removal of their stolonal bases and blended in distilled water at 4 °C. The whole-bodies homogenate was filtered in a spun glass mesh to remove large pieces of tissue and the filtrate was centrifuged twice in a Beckmann CS-6RK centrifuge at 1376× *g* during 30 min at 4 °C. Finally, the supernatants were recovered and freeze-dried.

Gel Filtration

The low molecular weight fraction from *Z. sociatus* was obtained by gel filtration chromatography of crude extract in Sephadex G50 matrix (Amersham Pharmacia Biotech, Uppsala, Sweden). Two grams (2 g) of crude extract were dissolved in 20 mL of 0.1 M ammonium acetate buffer (pH 6.7), centrifuged as described above and the supernatant was filtered through a 0.22 µm membrane (Merck Millipore, Billerica, MA, USA). The filtrate was applied to a chromatographic column (3.3 × 84 cm), packed with a Sephadex G50 matrix and previously equilibrated with the same ammonium acetate buffer. Chromatography was performed at a constant linear flow rate of 3.9 cm/h collecting fractions of 8.4 mL while monitoring elution through absorbance at 280 nm. Collected fractions were pooled in four major fractions and the third, usually the most prominent and containing low molecular weight compounds, was freeze-dried and used for experiments. The protein concentration was assessed using a bicinchoninic acid kit (Thermo Scientific, Rockford, IL, USA).

Mass Spectrometry Analysis

To evaluate the complexity of the fraction of interest MALDI-TOF/TOF (4800 Plus MALDI-TOF/TOF Analyzer; AB SCIEX, Framingham, MA, USA) spectra were obtained in positive linear mode from 700 Da to 12,000 Da and reflector positive mode 700–4000 Da

using the matrix α -cyano-4-hydroxycinnamic acid (α -CHCA) and also sequencing of peptide/metabolite in MS/MS mode. Samples were previously concentrated and cleaned according to the manufacturer's instructions on a micro C18 ZipTip column (Millipore, Bedford, MA, USA). Afterwards, some of these peptides were selected to cleave for partial sequencing.

Toxicity Test

Five groups of six OF-1 male mice were intraperitoneally (i.p.) administered with three doses of the fraction of interest: 150, 450 and 1350 μ g/kg in three 18–22 g OF-1 male mice each. Previous to the inoculation the fraction was dissolved in physiological saline (0.9% NaCl solution) and controls received a similar volume of vehicle alone. The percentage of mortality as a function of the dose was fitted to a Dose-Response curve to determine the mean Lethal Dose (LD₅₀) and the slope factor. Observations were done during the first hour post inoculation and the control and the surviving mice were monitored again after 24 h. Controls received a similar volume of saline and no dead was recorded. OF-1 mice were obtained from National Center for the Production of Laboratory Animals (CENPALAB), Cuba. Animal manipulation was performed according to the “International Guiding Principles for Biomedical Research Involving Animals” (Council for International Organizations of Medical Sciences, Geneva, Switzerland, 2012).

Data Analysis

Comparisons were performed by the non-parametric Mann-Whitney and Kruskal-Wallis (with a post-hoc Dunn) tests using GraphPad InStat version 3.00 (GraphPad Software, San Diego, CA, USA). Graphics were constructed using Origin version 7 (OriginLab Corporation, Northampton, MA, USA), as well as dose-response curve fitting.

2.5 Conclusion

The present study shows novel toxicological effects *in vivo* of a low molecular weight fraction from *Zoanthus sociatus* crude extract. MALDI-TOF mass spectra confirm that the low molecular weight fraction is composed by a mixture of non-peptides and peptides compounds between 700 and 6000 Da that caused signs of toxicity mainly related with cardiorrespiratory impairment and autonomic symptoms. The percentage of lethality showed a dose-dependent relation. Interestingly, our fraction of interest accelerated KCl-cardiac arrest, suggesting that the fraction could enhance the KCl-mediated cardiac dysfunction. Our report adds insights to the few studies related with this species and the potentialities of its low molecular weight compounds. However, further studies on this fraction should be performed into the composition of this fraction with the aim of elucidating the fraction components and possible mechanisms of action that support its toxicological properties.

2.6 Supplementary Information

The supplementary material is available on line at: <http://www.mdpi.com/1660-3397/11/8/2873/s1>

CHAPTER 3

Microcystin-LR Detected in a Low Molecular Weight Fraction from a Crude Extract of *Zoanthus sociatus*

Domínguez-Pérez D, Rodríguez A, Osorio H, Azevedo J, Castañeda O, Vasconcelos V, Antunes A: **Microcystin-LR Detected in a Low Molecular Weight Fraction from a Crude Extract of *Zoanthus sociatus***. *Toxins* 2017, **9**(3):89.

3.1 Abstract

Cnidarian constitutes a great source of bioactive compounds. However, research involving peptides from organisms belonging to the order Zoanthidea has received very little attention, contrasting to the numerous studies of the order Actiniaria, from which hundreds of toxic peptides and proteins have been reported. In this work, we performed a mass spectrometry analysis of a low molecular weight (LMW) fraction previously reported as lethal to mice. The low molecular weight (LMW) fraction was obtained by gel filtration of a *Zoanthus sociatus* (order Zoanthidea) crude extract with a Sephadex G-50, and then analyzed by matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrometry (MS) in positive ion reflector mode from m/z 700 to m/z 4000. Afterwards, some of the most intense and representative MS ions were fragmented by MS/MS with no significant results obtained by Protein Pilot protein identification software and the Mascot algorithm search. However, microcystin masses were detected by mass-matching against libraries of non-ribosomal peptide database (NORINE). Subsequent reversed-phase C18 HPLC (in isocratic elution mode) and mass spectrometry analyses corroborated the presence of the cyanotoxin Microcystin-LR (MC-LR). To the best of our knowledge, this finding constitutes the first report of MC-LR in *Z. sociatus*, and one of the few evidences of such cyanotoxin in cnidarians.

3.2 Introduction

Cnidarians represent promising sources of bioactive compounds, which can be of major pharmacological [330] and therapeutic interest [126, 145, 331]. However, many organisms belonging to the order Actiniaria (sea anemones), of the class Anthozoa, have been so far studied [14, 22, 27, 317]. In contrast, the order Zoanthidea has been scarcely explored for biologically active compounds, although some potent non-peptide toxins, such as palytoxin and its analogues, have been characterized from this order [100, 332, 333]. Additionally, a peptide exhibiting a reversible delay of tetrodotoxin (TTX)-sensitive sodium channel inactivation, was isolated and characterized from *Palythoa caribaeorum* exudate, but its sequence does not resemble any cnidarian toxin previously reported [334]. In another study the exudate of *P. caribaeorum* provoked reversible delay of the inactivation process of voltage-gated sodium channels ($\text{Na}_v1.7$), inhibited voltage-gated calcium ($\text{Ca}_v2.2$) and delayed rectifier (I_{DR}) currents of K_v channels from rat superior cervical ganglion (SCG) neurons [334]. The matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrometry (MS) analyses provided evidence of low molecular weight peptides involved in such effects on ion channels. It is also noteworthy the transcriptomic analysis on *Protopalythoa variabilis*, at which a transcript encoded a toxin called ShK/Aurelin-like that was toxic to zebrafish embryos [335].

Zoanthus sociatus also belongs to the order Zoanthidea. To date, few chemical studies have been reported from this organism, comprising the isolation and characterization of small non-proteinaceous molecules, such as the sterol zoanthosterol [336], as well as the norzoanthamines alkaloids and their analogues [337, 338]. In addition to these bioactive compounds, an antifilarial activity from the *Z. sociatus* chloroform/methanol extract was reported [322]. Moreover, only two reports on the toxicity of the crude extract and semi-pure chromatographic fractions have been published. The first study demonstrated that *Z. sociatus* components below 7 kDa are responsible for the inhibition of insulin secretion mediated by Ca^{2+} influx blockade in isolated rat β cells [114]. A further study showed the biological evaluation of a low molecular weight (LMW) Sephadex G-50-chromatographic fraction lethal to mice, presumably by cardiorespiratory arrest [339].

Information provided in previous studies about the molecular mass profile of the low molecular weight (LMW) fraction from *Z. sociatus* showed components in the m/z range from 700 to 6000 [114, 339]. Nonetheless, the mass spectrometry analysis of a LMW fraction obtained by the same methodology showed many m/z signals below 1000 [339]. Yet, the identification of the main components of the fraction remains to be done, thus, our main goal being the MALDI-TOF/TOF mass spectrometry analysis to characterize the most significant components of a LMW fraction from *Z. sociatus*. First, we performed the acquisition of a

MALDI-TOF/TOF MS spectrum of the LMW fraction followed by MS/MS fragmentation of the most significant signals. In general, the obtained data did not allow the identification of sea anemones-like peptide toxins or linear peptides using the Protein Pilot protein identification software or the Mascot algorithm search [200].

On the other hand, m/z signals from known cyanotoxins were detected by mass-matching against libraries of non-ribosomal peptide database (NORINE) [340, 341]. Indeed, within the microcystins masses detected, the m/z 995.53 matched the expected m/z signal for microcystin-LR (MC-LR). The presence of MC-LR was further demonstrated by high-resolution mass spectrometry analysis. While microcystins are recognized as freshwater cyanobacterial toxins, the occurrence of microcystins in marine ecosystems is not an isolated fact [342-346]. However, our findings constitute one of the few evidence of such cyanotoxin detected in cnidarians [347], and to the best of our knowledge the first report in *Z. sociatus*. Additionally, this work calls for further attention to the probabilities of water and food contamination by microcystins in tropical regions.

3.3 Results and Discussion

Gel Filtration in Sephadex G-50

The first purification step of the *Z. sociatus* crude extract was achieved by fractionation on Sephadex G-50. As previously described, the gel filtration chromatogram was divided into four fractions [19]. The resulting chromatogram showed the same elution profile of *Z. sociatus* crude extract, obtained by a similar purification protocol on HiLoad 16/20 Superdex 75 column [18]. The third fraction, called ZsG50-III, was then submitted to MALDI-TOF/TOF mass spectrometry analysis, given its lethal effect on mice as previously reported [19].

ZsG50-III MALDI-TOF/TOF Analysis

The selected fraction ZsG50-III was analyzed by MALDI-TOF/TOF MS analysis in positive ion reflector mode in the m/z range from 700 to 4000. The most intense signal detected in the MALDI-TOF/TOF MS analysis corresponded to m/z values below 2000, resulting in the detection of 142 signals, ranging from m/z 703.93 to 1336.96. The highest intensity signals are shown in Figure 3.1. Afterwards, some of the most intense and representative MS ions corresponding to m/z 876.98, 861.01, 1050.04, and 1066.00 were fragmented by MS/MS (Figures 3.2 and 3.3). Then, all the data generated by the MALDI-TOF/TOF procedure were submitted to search with the Mascot (Matrix-Science, London, UK) algorithm against UniProtKB protein sequence database [204], specifically in the Metazoan and Cnidaria section. Additionally, the spectra were also analyzed with the Protein Pilot protein identification software v4.5 (AB SCIEX). No significant result was obtained by

any of the mentioned methods. The fragmentation pattern of the selected ions showed scarce peaks. In fact, the spectra do not resemble those of linear peptides, but seem to be related to cyclo-peptides fragmentation.

On the other hand, cyanotoxins-related masses were detected by peak list mass-matching against libraries of non-ribosomal peptide database (NORINE). Indeed, some of the most intense signals were related to microcystins. The mass m/z 1066.00 matched a reported microcystin [9-acetyl-Adda5]-MC-RR [348]. Furthermore, the MS/MS analyses of m/z 1066.00 showed a high intensity signal at m/z 876.94 (Figure 3.2). Besides, in the MS/MS spectra of m/z 1066.00 and m/z 876.94 signals at m/z 265.95 and 627.05, were detected, respectively (Figure 3.2). These masses were previously found in the LC/ESI-Q-ToF-MS/MS spectrum of the microcystin [9-acetyl-Adda5]-MC-RR [348]. Other important diagnostic signals such as m/z 135 corresponding to 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-deca-4,6-dienoic acid (Adda) fragment, or m/z 163 corresponding to acetyl-Adda side chain were absent or could not be detected because of their low intensity.

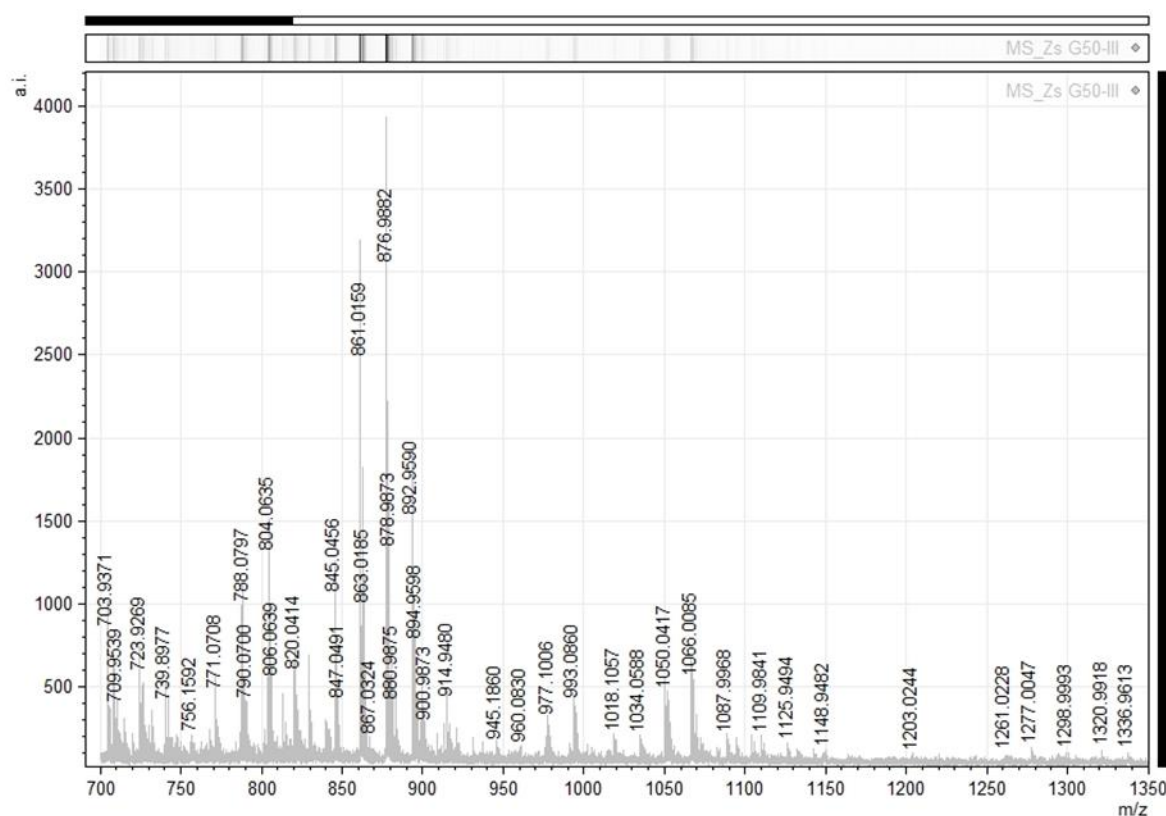
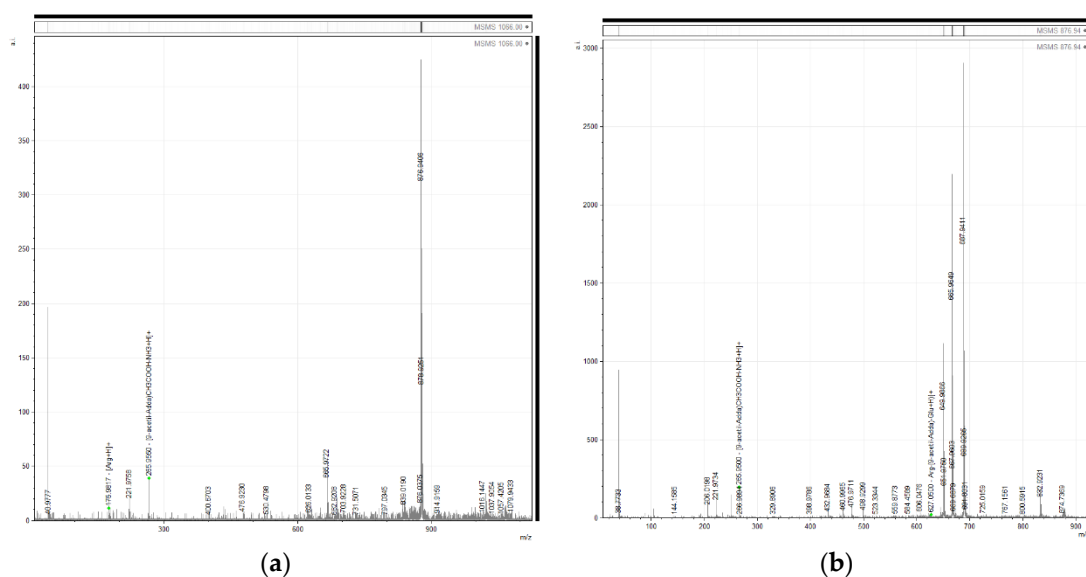


Figure 3.1. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF) mass spectrum of the Sephadex G-50 fraction called ZsG50-III. The mass spectrum shows absolute ion intensity (a.i.) versus mass-to-charge ratio (m/z) of ZsG50-III components for the mass range m/z 700 to 1350.



fragmentation pattern of m/z 1050.04 and m/z 861.01 showed common signals (Figure 3.3). Contrasting with the analysis of m/z 1066.00, it was not possible to match the fragments of m/z 1050 with known signals from fragmentation of MC-(H₄) YR [349]. The diagnostic fragments annotation was hindered because of the limited information from the MC-(H₄) YR m/z 1050 spectrum. Moreover, in this case, the spectrum also lacked the microcystins diagnostic signals like Adda fragments m/z 135 and m/z 163 [348].

Moreover, some microcystin-related m/z signals were detected in the MALDI-TOF analysis of ZsG50-III within the m/z range 960–1050 (Figure 3.4). These signals are m/z 960.0830, microcystin YA (MC-YA) [351]; m/z 981.0593, demethylated variant of MC-LR ([DMAdda5]microcystin-LR) [350]; m/z 995.0931, MC-LR (microcystin-LR) [352]; m/z 1002.1186, [D-Asp3.Ser7]microcystin-E(OMe)E(OMe) [353] or microcystin LY (MC-LY) [354]; m/z 1010.0557, [D-Asp3.Dha7]microcystin-RR [355]; m/z 1024.0260, [D-Asp3]microcystin-RR [356]; m/z 1029.1227, microcystin FR (MC-FR) [357]; m/z 1039.0340, [D-Ser1. ADMAdda5]microcystin-LR [358]; m/z 1045.0357, microcystin YR (MC-YR) [352] or [Dha7]microcystin-HtyR [359]. However, these m/z signals were not successfully sequenced by MS/MS and, subsequently, analytical steps were conducted to corroborate the presence of microcystins.

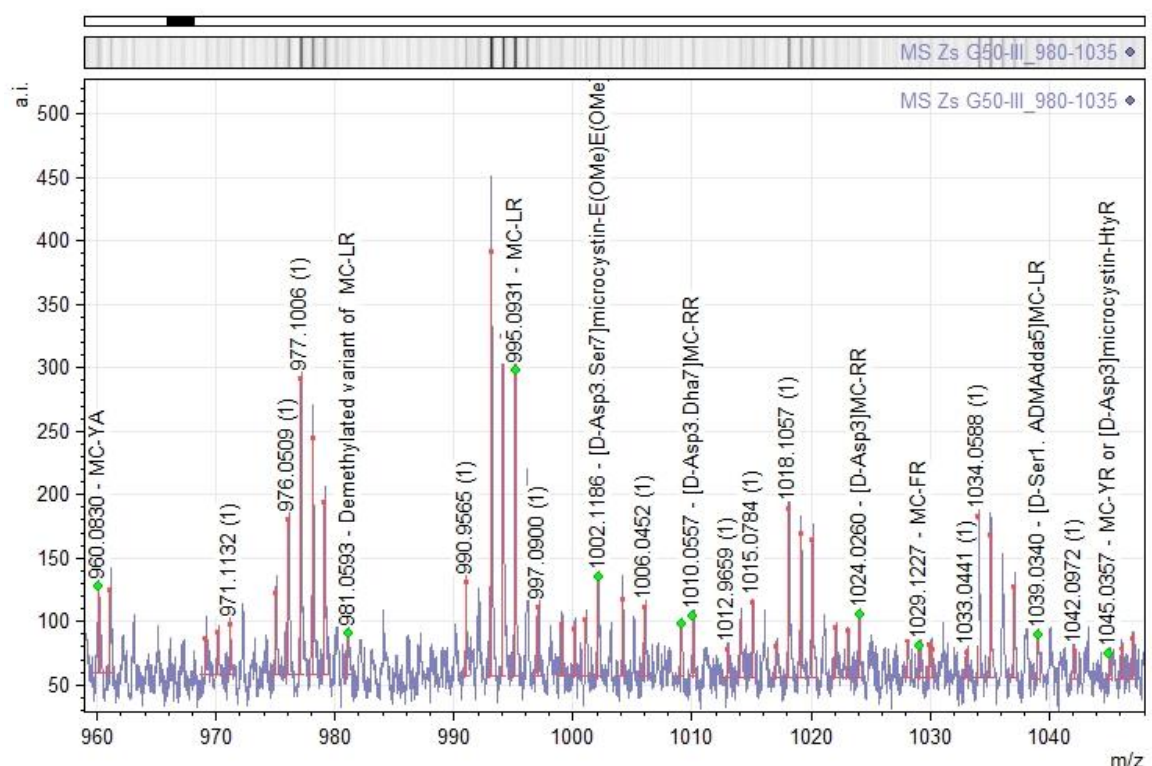


Figure 3.4. The figure shows absolute intensity (a.i) versus mass-to-charge ratio (m/z) in the m/z range 960–1050 from the MS analysis of fraction ZsG50-III. Some masses are highlighted as annotations, considering similar m/z matches as possible microcystins.

ZsG50-III Reserved-Phase Chromatographic Analysis

RP-HPLC Analytical Profile

The detection of some microcystin-related m/z signals, prompted the performance of an analytical chromatographic step to ensure the identification of microcystins. Taking into account that the signal m/z 995.09 matched MC-LR, as the most commonly known and widely distributed microcystins, efforts were focused on confirming the presence of MC-LR in the sample. Therefore, the fraction ZsG50-III was first analysed by RP-HPLC in analytical mode to obtain preliminary information on the sample complexity. The analytical chromatogram showed what we considered as nine peaks of moderate intensity, where the highest relative absorbance was produced by the less retained compounds (Figure 3.5). Additionally, some of the peaks detected in the analytical chromatogram of the fraction ZsG50-III produced UV spectra with a λ_{max} at 220 and 274 nm, while the other peaks' UV maximum-absorbance were around 233, 260, and 300 nm (Supplementary Figure S1), close to the maximum absorbance of microcystins.

In addition, a chromatogram was obtained with a commercial standard of MC-LR in analytical mode in the same conditions as mentioned for fraction ZsG50-III (Figure 3.5). The elution time for the MC-LR standard was 8.8 min, whereas in the analytical mode three closely-eluting peaks at 7.3, 8.0, and 8.6 min were detected. Unlike MC-LR, such peaks do not show maximum absorbance at 238 nm (Supplementary Figure S1). This behaviour could be explained by insufficient separation of the components.

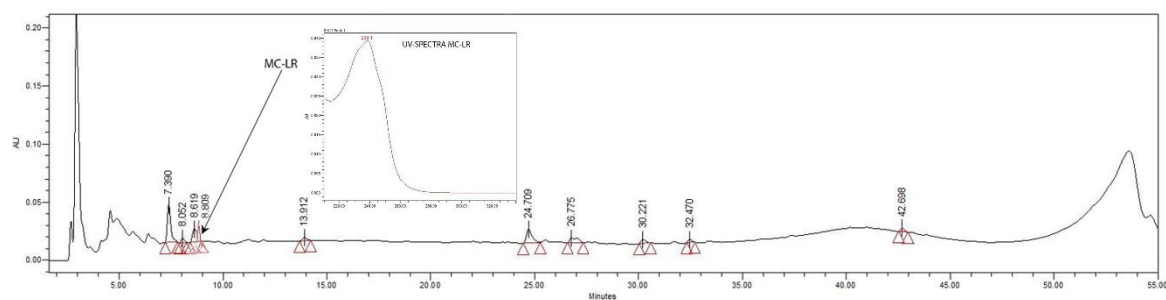


Figure 3.5. Analytical profile of fraction ZsG50-III, obtained by RP-HPLC. Gradient elution started from 1% to 99% MeOH over 55 min. The injected volume was 10 μL at a concentration of 1 mg/mL. The PDA range was 210–400 nm, with fixed wavelengths at 220 nm and 280 nm. The peak of the commercial standard of MC-LR (in red line) eluted at a retention time of 8.8 min. The injected volume was 10 μL at a concentration of 1 mg/mL. The UV spectrum of MC-LR shows maximum absorbance at 238 nm, typical of microcystins.

ZsG50-III RP-HPLC Semi-Preparative Assays

Afterwards, a semi-preparative RP-HPLC run (Figure 3.6) was carried out to separate those components with retention time like that of the commercial MC-LR chromatogram (Figure 3.5). Unlike microcystins, the UV maximum absorbance for the first peak was 264 nm, whereas the second peak showed a UV maximum absorbance at 220–274 nm (Supplementary Figure S2), resembling a typical UV spectra of peptides [360–363]. In contrast to Liquid chromatography combining multi-stage mass spectrometry (MC-LC/MS), UV-based methods do not provide unequivocal identification of known, unexpected and/or trace levels of microcystins [364].

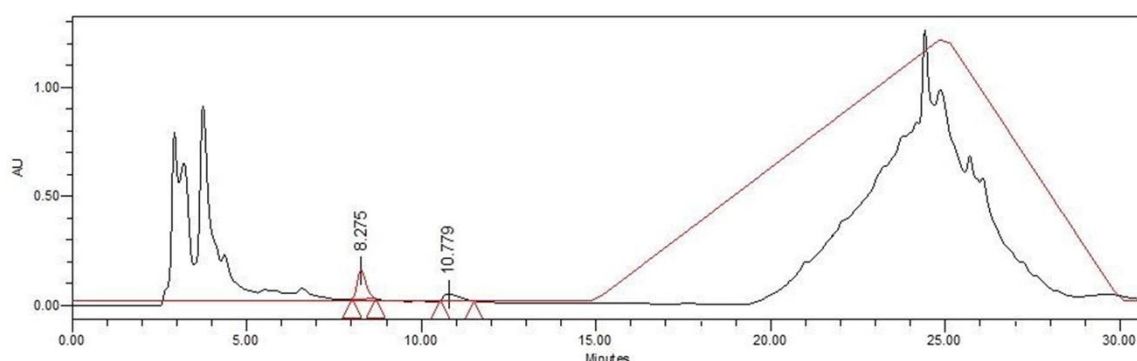


Figure 3.6. Chromatogram of fraction ZsG50-III subjected to reversed-phase C18 HPLC in isocratic mode with 10% MeOH using a 0.1% TFA/H₂O/MeOH elution system. After 15 min the solvent was increased from 10% to 99% MeOH over 30 min. The chromatogram obtained by the RP-HPLC procedure shows relative absorbance at 220–280 nm of two peaks and their respective retention times of 8.2 min for peak 1 and 10.7 min for peak 2.

Nonetheless, the UV spectra of the peaks were not clearly related to MC-LR. Other cyclopeptides with similar elution time, maximum UV absorbance at 220–274 nm and m/z 995 were previously reported as cyanopeptolin [365]. However, the isomerization [366] and co-elution with another congeners, like cyanopeptolin [365, 367] and microginin [368], can produce unusual UV spectra. The co-occurrence of cyanotoxins with other congeners is a common fact supporting the possibility of components co-elution. It is noteworthy the presence of putative co-eluting components detected by MS of the peak 2 (Supplementary Figure S3). Unfortunately, the amount of the fraction was limited, and further purification steps based on the gradient elution with phosphoric acid as the mobile phase was not possible to perform as previously reported [368].

Identification of MC-LR by MALDI-TOF/TOF Analysis

The two peaks obtained by RP-HPLC were first submitted to MALDI-TOF MS scan resulting in the detection of m/z values common to both peaks, such as m/z 795.3, 912.3, 926.4, 930.3; 944.4; 995.5, 1007.4, 1011.5, and 1029.5 (Figure 3.7). Nonetheless, some

differences were detected in the peak list and signal intensity, among which the m/z 765.33 showed the highest intensity in peak 1, whereas 995.53 was the most intense m/z signal in peak 2. The mentioned m/z 995.53 was successfully sequenced via MS/MS analysis of peak 2 (Figure 3.8) and the resulting spectrum matched the m/z values of some fragments produced by MS/MS analysis of MC-LR (Figure 3.8). Some fragments from the MS/MS of m/z 995.53 were successfully annotated, including m/z 135, as a diagnostic characteristic of microcystins. Some of the fragments detected were annotated, such as m/z 70.07 [Arg-related ion], m/z 135.09 [Adda fragment], m/z 375.15 [Adda-fragment + Glu + H], 599.27 [Arg + Adda + Glu + H], and 861.33 [M-134]. Others m/z signals, like 1224.59, 1471.68, 1731.82, and 1866.87, corresponding to co-eluting components to the MC-LR (m/z 995.53), were also successfully fragmented on peak 2, (Supplementary Figure S3). Although, it was not possible to identify them, the m/z signal 1731.82 and 1866.87 resembled microviridins previously described [369].

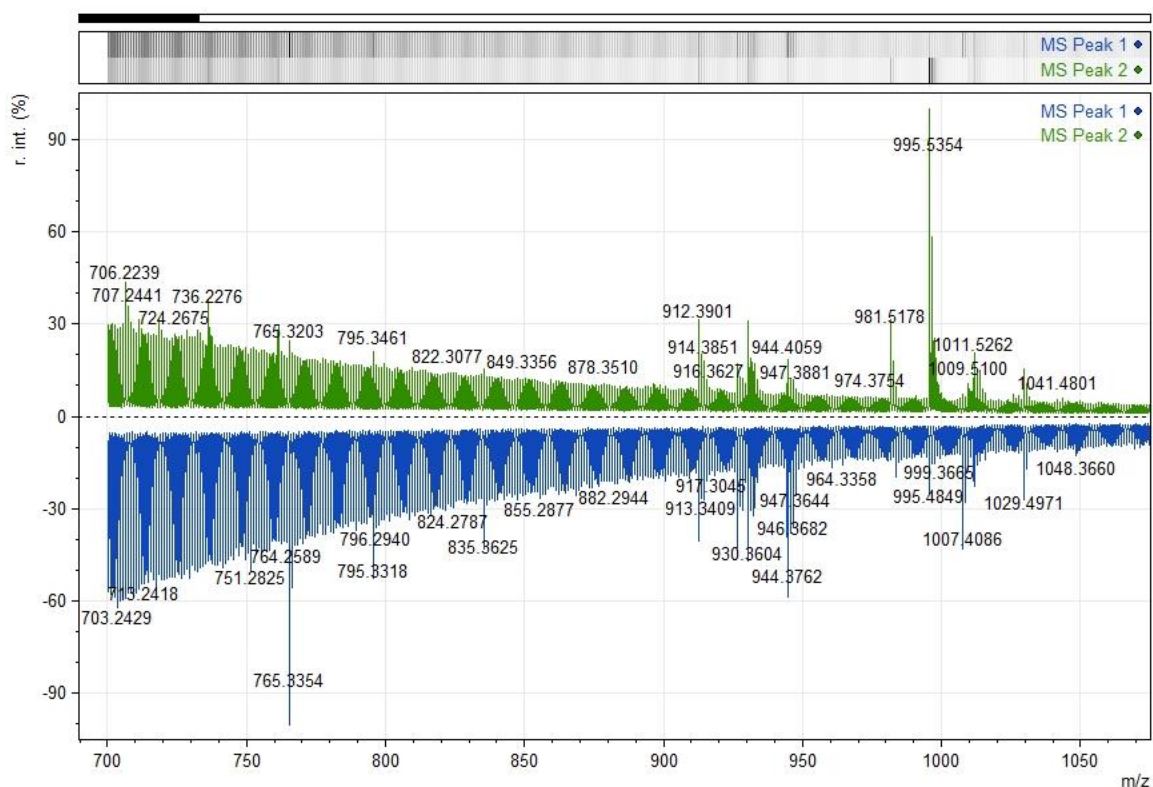


Figure 3.7. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrum of the two RP-HPLC peaks. The composite figure shows relative intensity (r.int) versus mass-to-charge ratio (m/z) in the range of m/z 700–1075 of peak 1 (blue) peak 2 (green).

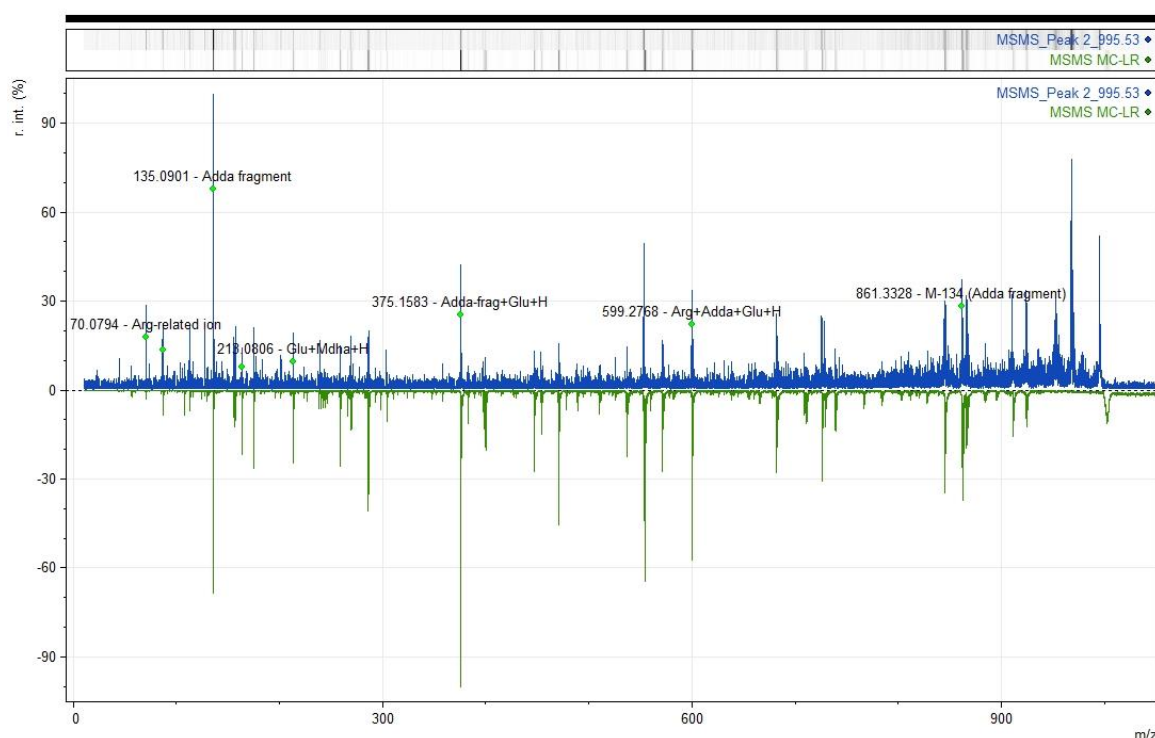


Figure 3.8. MS/MS of MC-LR. The figure shows relative intensity (r.int) versus mass-to-charge ratio (m/z) or the spectrum generated by MALDI-TOF/TOF of the signal m/z 995.53 from peak 2 (color blue, top panel) and the MS/MS of 1 μ g of an MC-LR commercial standard (blue, bottom panel). Note similarities in the fragmentation pattern in both spectra; the most relevant ions in the identification of the MC-LR are annotated in the top panel.

Putative Origins of Microcystins in the ZsG50-III Fraction

Cyanobacteria also known as blue-green algae are a widely distributed group of photosynthetic prokaryotic organisms [350, 370, 371]. Some genera like *Microcystis*, *Planktothrix*, *Anabaena*, *Nostoc*, and *Nodularia* can produce diverse toxins called cyanotoxins [350]. Within them, the microcystins mostly produced by *Microcystis aeruginosa*, comprise more than 100 variants of such cyanotoxin [372]. The microcystins are synthesized via a non-ribosomal pathway, where peptide synthetases (NRPS) and polyketide synthases (PKS) play an important role [373]. The general structural of these cyclic heptapeptides are: cyclo [D-Ala(1)–L-X(2)–D-MeAsp(3)–L-Z(4)–Adda(5)–D-Glu(6)–Mdha(7)], where L-X and L-Z in position 2 and 4 of the ring are variable L-amino acids, D-MeAsp is a non-proteogenic aminoacid D-erythro- β -methylaspartic acid, Mdha is N-methyldehydroalanine [374]. The Adda, that is (2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid, is a diagnostic characteristic of microcystins. Each microcystin is named depending on the identity of X and Z amino acids [348].

Cyanobacteria toxins are more common in world-wide freshwater ecosystems [375, 376], although cyanobacteria are widespread in estuarine and marine systems, as well [377, 378]. However, there are reports on the occurrence of microcystins in marine ecosystems

producing hepatic necrosis, haemorrhage, and sudden death in marine mammals [342, 379]. The most significant report produced by microcystins intoxication produced the death of 21 southern sea otters [342]. The land-sea flow with trophic transfer through the marine invertebrate-producing bioaccumulation process was demonstrated. It is noteworthy that our specimens belong to the invertebrate *Z. sociatus*, sampled at the Quibú river mouth, in Havana City. Moreover, this river passes through the city and is likely to be eutrophized, providing the proper conditions for the occurrence of microcystins.

Additionally, the confirmation of microcystin-contaminated freshwater outflows to the ocean is not an isolated and sporadic event [342-346]. There is also evidence of microcystins in some marine animals, like copepods, corals, and fish [347, 380, 381]. In addition, bioaccumulation of microcystins have been detected on fresh and saltwater mussels [382, 383], farmed crustaceans [384, 385], fishes [381], and probably in humans [386]. In our case, the cyanobacteria-producing toxins should come from the river flow to the marine zone, occurring as a guest of *Z. sociatus* until being fed. Then microcystins can be incorporated into the *Z. sociatus* tissues. Therefore, we suggest that the most probable source of the toxins detected in our sample is an external producer and not a product of the zoanthid *Z. sociatus*. However, the hypothesis of *Z. sociatus* as responsible for the production of microcystin should not be completely discharged, since an analogue of microcystins, called motuporin (analogue of Nodularin-R), was first isolated and described in the marine sponge *Theonella swinhoei* [387].

Taking into account the high toxicity of microcystins on mammals, specifically in mice [388], the presence of microcystins in the studied fraction could be related to the lethal effects on mice previously reported of a LMW fraction from *Z. sociatus* [339]. In fact, MC-LR and its congeners are highly toxic in mice (lethal dose: LD₅₀ = 50 µg/kg) when administered intraperitoneally [389]. Unlike ZsG50-III, mostly of the known MC-LR toxic effects are hepatotoxic. In addition, the mentioned studies lack of enough elements to know if MC-LR was tested in an acute toxicity assay producing death in a short time. Nonetheless, the possibilities of MC-LR be involved in cardiotoxic effects could not be discharged, since it is capable of modulating Ca²⁺ channels. Indeed, the presence of microcystins on the fraction ZsG50-III could also be related to the effects on insulin secretion mediated by Ca²⁺ influx blockade in isolated rat β cells [114]. In fact, exposure to MC-LR for 72 h suppresses cell viability, disturbs glucose-stimulated insulin secretion, and decreases the expression of insulin protein [390]. Although it is not clear if the presence of microcystins are responsible for the mentioned effects, they are probably largely involved.

3.4 Conclusions

The present work represents one of the few attempts to identify new peptide toxins from the zoanthid *Z. sociatus*. However, the mass spectrometry analysis of a fraction from *Z. sociatus* resulted in the detection of masses below 2000 Da, among which none sea anemones-like toxins were detected with the Protein Pilot protein identification software and Mascot algorithm search. Nonetheless, some related microcystins masses were detected within the most intense signals generated by the mass spectrometry analysis. One of them, m/z 995.53 resulted in a highly similar fragmentation pattern than that of MC-LR standard. Most of the fragment produced by the ionization of the m/z 995.53 were successfully annotated, including m/z 135.09 [Adda fragment], as a diagnostic characteristic of microcystins. The combination of the evidences detected led to confirm the presence of MC-LR in the fraction ZsG50-III. To date, this finding constitutes one of the few pieces of evidence of such cyanotoxin being detected in cnidarians, and the first report in the zoanthid *Z. sociatus*. In the cases of m/z 1066.00 and 1050.04, we considered that the information provided by the MS/MS spectra is insufficient to identify these signals as coming from microcystin [9-acetyl-Adda5]-MC-RR and MC-(H₄) YR, respectively. Additionally, the presence of microcystins in fishing areas with similar conditions to that of the *Z. sociatus* sampling area remain unexplored in Cuba. Considering the mentioned evidence, the occurrence of microcystins in fishes should be of high probability around the sampling place. Therefore, these findings clearly suggest the underestimated risk of intoxication by microcystins occurring in water and food in a tropical region.

3.5 Materials and Methods

Preparation of Crude Extract and Gel Filtration on Sephadex G-50

The preparation of the *Z. sociatus* crude extract was performed from wild-caught colonies sampled at Quibú River mouth, in Havana, Cuba. The gel filtration chromatographic separation on Sephadex G-50 were performed as previously described [339]. Briefly, *Z. sociatus* specimens were blended after removing their stolonal bases. The whole-body homogenate was filtered through a spun glass mesh, and the filtrate was centrifuged twice in a Beckmann CS-6RK centrifuge at $1376\times g$ for 30 min at 4 °C. The supernatant was freeze-dried overnight and then submitted to fractionation on a Sephadex G-50. The most prominent fraction of the chromatogram, called ZsG50-III was freeze-dried and stored for subsequent experiments.

Mass Spectrometry Analysis and Database Search

Mass spectrometry analysis was performed by MALDI-TOF/TOF (4800 Plus MALDI-TOF/TOF Analyzer; AB SCIEX, Framingham, MA, USA). Mass spectra were analysed with

the Data Explorer software (v3.7, build 126, AB SCIEX, Framingham, MA, USA). The samples obtained in the semi-preparative HPLC-PDA separation, corresponding to peaks 2, 3, 4, 5, and 6, were co-crystallized at room temperature with the matrix α -cyano-4-hydroxycinnamic acid, α -CHCA at 10 mg/mL (50% acetonitrile, and 0.1% trifluoroacetic acid (TFA)) in a MALDI target plate. Samples were previously concentrated and cleaned according to the manufacturer's instructions on a micro C18 column (ZipTip, Millipore, Bedford, MA, USA). The mass spectra were acquired in positive ion reflector mode from m/z 700 to 4000. Afterwards, some of these peptides peaks were selected for MS/MS peptide sequencing.

The masses $[M + H]^+$ and the corresponding native intensity of HPLC-PDA peaks obtained from MS scan were exported to ASCII format for further analysis with the mMass 5.5.0 software [391, 392]. Firstly, Peak Lists was achieved using the following criteria: peak's signal-to-noise ratio threshold (S/N) 2.0; absolute intensity threshold (a.i) 10.0 (peak's native intensity without baseline correction); relative intensity threshold (r.int.) 0.3 (peak's relative intensity, in percentage of the most intense peak); and picking height of 75. Additionally, baseline correction, smoothing, and deisotoping tools were also applied to all MS fixed at default settings. The matrix α -CHCA monoisotopic masses $[M + H]^+$ acquired in the same MS mass window were used as control for calibration. The resulting peak lists were compared to each other using a tolerance of 0.1 Da and different masses were determined. Finally, all different candidates' masses $[M + H]^+$ were corroborated in each MS (each MS-peak-spectrum against MS-matrix-spectrum) in both mode Autoscale and Normalize View.

The data generated by MALDI-TOF/TOF procedure were submitted to search with the Mascot (Matrix-Science, London, UK) algorithm. The protein identification was carried out from UniProtKB protein sequence database [204] in the Metazoa and Cnidaria section. The Mascot search parameters were fixed as follows: up to two maximum trypsin missed cleavages, mass tolerance of 50 ppm, cysteine carbamidomethylation (fixed modification), methionine oxidation (variable modification), and a charge state of +1. The same data mentioned were also analysed by Protein Pilot protein identification software v4.5 (AB SCIEX). The protein identification was carried out from UniProtKB protein sequence database in the Metazoa and Cnidaria section. Only those proteins matched with scores at a 95% confidence level were considered as significant hits.

Moreover, all spectra generated for each peak were also analyzed by a mass-matching tool provided by the mMass software v5.5.0 [391, 392]. In this case, the analysis was carried out within the internal monomer mMass library and also against the imported non-ribosomal peptides library (NORINE peptide database, available in the mMass homepage at the download section <http://www.mmass.org/download/>), adapted from the original NORINE database [340]. The matched masses were then examined, and their

spectra were analyzed. The fragmentation pattern was compared to a standard or available reference, in the case of those that were of interest.

MALDI-TOF/TOF Raw Data Availability

The output files generated by MALDI-TOF/TOF analyses used in this work are provided in ASCII format as Supplementary Raw Dataset S1. Within the package are included the following folders containing the respective mass spectra (MS/MS): Matrix (MALDI TOF/TOF MS of matrix used as control; MS_ZsG50-III (MS of studied fraction ZsG50-III); MS/MS_1066_&_876 (MS/MS of the signals m/z 1066.00 and 876.94 from the MS of fraction ZsG50-III); MS/MS_1050_&_861 (MS/MS of the signals m/z 1050.04 and 861.01 from the MS of fraction ZsG50-III); Peak 1 (MS of Peak 1 was obtained by RP-PHLC from ZsG50-III fraction); Peak 2 (MS of Peak 2 was obtained by RP-PHLC from the ZsG50-III fraction, and the MS/MS of the signals m/z 995.53, 1224.59, 1471.68, 1731.82, and 18,866.87), and MS_&_MS/MS_MC-LR (MS of the commercial standard of MC-LR, and MS/MS of the signals m/z 995.53 and 981.50 were obtained from the MS of MC-LR).

Reserved-Phase Chromatographic Analysis

All solvents used in RP-HPLC analyses were of high-purity chromatography grade (LiChrosolv, Merck, Darmstadt, Germany). Aqueous solutions were prepared with ultrapure water supplied from a Millipore water purification system ($0.0054 \mu\text{S}\cdot\text{cm}^{-1}$). Trifluoroacetic acid (TFA) was of spectrophotometric grade 99%. The chromatographic system used for RP-HPLC separations was a Waters Alliance e2695 HPLC coupled with a photodiode array (PDA) 2998 and an automatic fraction collector. Empower Two Chromatography Data Software was used for calculation and reporting peak information. All HPLC solvents were filtered (Pall GH Polypro 47 mm, $0.2 \mu\text{m}$) and degassed by an ultrasonic bath.

The analytical profile was obtained on a Merck Lichrospher RP-18 endcapped reversed-phase column ($250 \text{ mm} \times 4.6 \text{ mm i.d.}$, $5 \mu\text{m}$) equipped with a guard column ($4 \times 4 \text{ mm}$, $5 \mu\text{m}$), both kept at 40°C . The PDA wavelength range was 210–800 nm with fixed values at 220 and 280 nm. The solvent system consisted of MilliQ water (H_2O) and methanol (MeOH), both acidified with 0.1% TFA. A linear gradient from 1% to 99% MeOH was applied in the analytical run to explore the complexity of the sample and the elution time of the commercial standard of Microcystin-LR (DHI Water and Environment, Denmark; Batch No: MCLR-111).

The purification step was performed using a Phenomenex Luna RP-18 ($25 \text{ cm} \times 10 \text{ mm}$, $10 \mu\text{m}$) chromatographic column kept at 35°C . The PDA range was 210–800 nm with fixed wavelengths of 220 and 280 nm. For the isocratic elution, 0.1% TFA in 10% MeOH was used as eluent at a flow rate of 2.5 mL/min . Eventually, the solvent of the mobile phase was increased to 50% MeOH. The concentration of total-proteins estimated was 1 mg/mL , and

injected in a final volume of 200 μ L. The fractions were automatically collected following the PDA signal. The fractions of interest were dried up by speed vac and kept at -20°C for subsequent analysis.

3.6 Supplementary Materials

The following are available online at www.mdpi.com/2072-6651/9/3/89/s1 and/or goo.gl/OsrgQz. Figure S1: The figure shows the UV spectra of peaks obtained from the RP-HPLC analytical profile of fraction ZsG50-III, Figure S2: The figure shows the UV spectra of two peaks obtained from the RP-HPLC of the fraction ZsG50-III in a semi-preparative mode, Figure S3: Matrix assisted laser desorption/ionization time-of-flight/time-of-flight(MALDI-TOF/TOF) mass spectra (MS) of the RP-HPLC Peak 2 and Supplementary Raw Dataset S1. The Raw Dataset is available via <http://creativecommons.org/licenses/by/4.0/> at https://www.researchgate.net/publication/316018751_Supplementary_Material/data/58edb012458515c4aa50f738/toxins-09-00089-s001.zip

CHAPTER 4

Proteomic analyses of the unexplored sea anemone *Bunodactis verrucosa*

Dany Domínguez-Pérez^{1,2†}; Alexandre Campos^{12†}; Armando Alexei Rodríguez³, Maria V Turkina⁴; Tiago Ribeiro¹², Hugo Osorio^{5,6,7}, Vítor Vasconcelos^{1,2} and Agostinho Antunes^{1,2*}

¹ CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, Porto 4450-208, Portugal

² Biology Department, Faculty of Sciences, University of Porto, Rua do Campo Alegre, s/n, Porto 4169-007, Portugal

³ Department of Experimental and Clinical Peptide Chemistry, Hanover Medical School (MHH), Feodor-Lynen-Straße 31, D-30625 Hannover, Germany

⁴ Department of Clinical and Experimental Medicine, Linköping University, SE-581 85, Linköping, Sweden

⁵ i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen, 208

⁶ Ipatimup, Institute of Molecular Pathology and Immunology of the University of Porto, Rua Júlio Amaral de Carvalho, 45, 4200-135 Porto, Portugal

⁷ Department of Pathology and Oncology, Faculty of Medicine, University of Porto, Al. Prof. Hernâni Monteiro, 4200-319 Porto, Portugal

†authors contributed equally to this work

*To whom correspondence should be addressed: aantunes@ciimar.up.pt

Keywords: shotgun proteomics, cnidarian, sea anemones, proteins, toxins, MALDI-TOF/TOF, Two-dimensional gel electrophoresis

This chapter will be submitted for publication

4.1 Abstract

Cnidarian toxic products, particularly peptide toxins, constitute a promising target for biomedicine research. Indeed, cnidarians are considered as the largest phylum of generally toxic animals. However, research on peptides and toxins of sea anemones is still limited. Moreover, most of the toxins from sea anemones have been discovered by classical purification approaches. Recently, high-throughput methodologies have been used for this purpose but in other Phyla. Hence, the present work was focused on the proteomic analyses of whole-body extract from the unexplored sea anemone *Bunodactis verrucosa*. The proteomic analyses applied were based on two-dimensional gel electrophoresis combined with MALDI-TOF/TOF and shotgun proteomic carried out by nano-LC coupled to a hybrid Ion trap mass spectrometer (LTQ Orbitrap) and the obtained raw data was searched against custom protein databases using MaxQuant software. In total, 413 proteins were identified by shotgun proteomics approaches, gel-based approaches and shotgun proteomics analyses. The most significant GO terms for each of the three major GO categories were “metabolic process”, “binding” and “cell parts”. On the other hand, eight proteins identified from gel-based analyses, were mainly involved in Glycolysis/Gluconeogenesis pathway, antioxidants activities and RNA degradation. In addition, KEGG analyses of all proteins identified revealed Purine metabolism, Thiamine metabolism and Biosynthesis of antibiotics as the most relevant pathways. Moreover, some toxins including metalloproteinases and neurotoxins were also successfully identified. The mechanism of action of such toxins in prey catching and feeding is proposed, which seemingly act synergically. The present work provides the first proteome map of the sea anemone *B. verrucosa*. To our knowledge, the results presented constitute new findings for this sea anemone.

4.2 Introduction

The cnidarians represent the largest source of bioactive compounds, as candidates for pharmacological tools [330] and even new drugs for therapeutic treatments [126, 145, 331]. Unlike toxin from terrestrial animals, cnidarian venoms have not received as much scientific attention [295]. Each one of around 11 000 living species [292] possess nematocysts [313], which is the organ specialized in the production, discharge and inoculation of toxins [156]. Hence, the toxic feature can be theoretically ascribed to all the member of this Phylum, since nematocysts is the only of the three categories of cnidae found in all cnidarians [156]. However, without including components of the venom described at the transcriptomic level, only about 250 compounds had been reported until 2012 [22], although this figure has not increased significantly at the proteomic level in the last five years. The composition of cnidarians is composed mainly by peptides, proteins, enzymes, protease inhibitors and non-proteinaceous substances [22].

Most of the known toxins from cnidarians belong to the Order Actiniaria, Class Anthozoa (sea anemones) [14, 22, 155]. Among sea anemones, there were recognized to date around 200 non-redundant proteinaceous toxins, including protein and peptides [317, 393]. Besides, other 69 new toxins were revealed by transcriptomic-based analyses, although, 627 candidates were additionally proposed comprising 15 putative neurotoxins [394] and 612 candidate toxin-like transcripts from other venomous taxa [395]. In general, peptides toxins from sea anemones can be classified as cytolytins, protease inhibitors or ion channel toxins (neurotoxins), mainly voltage-gated sodium (Na_v) channel toxins and voltage-gated (K_v) potassium channel toxins [14, 22, 27, 58, 396]. Sea anemones have a great potential as a source of peptide/protein toxins within cnidarians, partly because their toxins are considerably stable compared to other cnidarian toxins (e.g. jellyfishes), yet only a limited number of sea anemone have been examined for peptide/protein toxins [14], although more than 1000 species have been recorded [157]. Thus, sea anemones represent a relatively unexplored potential source of bioactive/therapeutic compounds.

Particularly, *Bunodactis verrucosa* is one of the most common species of sea anemones in the intertidal zone of Portugal coast [294], but its proteome, including peptide toxins, remains unexplored. The main goal of this work is to address a general proteomic characterization of the whole-body extract from the sea anemone *B. verrucosa*. This species also known as *Aulactinia verrucosa* occurs at the northeastern Atlantic Ocean, the North Sea and the Mediterranean Sea [397]. The specimens were sampled at Praia da Memória, Porto, Portugal. After sample preparation, the extract containing the digested protein was subjected to shotgun analysis. The resulting raw data from LC-MS/MS were searched against some protein databases using MaxQuant software. At this end several proteins were successfully

identified, including several potential toxins. Until now, just a few chemical studies have been reported from *B. verrucosa*. In fact, to the best of our knowledge, this study provides the first proteomic profile of this species. Most of the proteins identified constitute first reports for this species.

4.3 Materials and Methods

Protein extraction

Specimens of *B. verrucosa* were sampled at Praia da Memória, Porto, Portugal (Lat/Long WGS84; N 41° 13.84032' W 8° 43.27464'). Then whole animal bodies (four specimens) were kept at -80°C, freeze dried and subsequently homogenized in a blender until obtaining a dry powder. Lyophilized material of *B. verrucosa* (0.1g) was mixed with 500 µl Tris-HCl (40mM), MgCl₂ (5mM), Dithiothreitol (DTT) (1 mM), protease inhibitors (87785, ThermoScientific) at pH 8.0, (buffer 1) in vortex (2 * 30 sec.). The mixture was centrifuged at 16000 x g, during 20 min at 4°C. The supernatant (soluble protein fraction, SF) was stored at -20°C and the pellet was homogenized (overnight, at 4°C) with 500 µl urea (7 M), thiourea (2 M), CHAPS (4%, w/v), dithiothreitol (65mM) and ampholytes (0.8%, v/v), at pH 4–7 in vortex (2 * 30 sec.). The homogenate was centrifuged at 16000 x g during 20 min at 4°C, and the supernatant (insoluble protein fraction, IF) collected and stored at -20°C. Total protein concentration was estimated according to the Bradford method [168].

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis (2DE) was performed as described previously [398]. Duplicate IF and SF (~ 400 µg of protein) were diluted to 300 µL in sample buffer (SB, (x mM sodium borate, at pH 8.0) and loaded onto 17 cm, pH 4–7 immobiline dry strips (Bio-Rad) with active hydration (50 Volt) for 12 h. Proteins were separated by isoelectric focusing (IEF) in a Protean IEF cell (Bio-Rad) with the following program: step 1, 15 min at 250 V; step 2, 3 h voltage gradient to 10 000 V (linear ramp); step 3, 10 000 V until achieving 60 000 V/ h (linear ramp). Second-dimension SDS-PAGE was performed in a Hoefer SE 900 vertical slab electrophoresis system (Hoefer, Holliston, MA, USA), with 12 % (w/v) acrylamide gels, at 480 mA and 20 °C. After electrophoresis run the gels were stained with colloidal Coomassie blue G-250 [399]. The 2DE protein profiles were analyzed by gel scanning with a GS-800 calibrated densitometer (Bio-Rad) and the PDQuest 2D analysis software (Bio-Rad) as described previously [398]. Protein spots detected by this procedure were excised from the gels for subsequent identification.

MALDI-TOF MS analysis

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF/TOF) mass spectrometry (MS) measurements were performed to identify protein spots from 2DE gels.

Protein spots were washed, destained, reduced, alkylated, and digested with trypsin following the procedure described by Osório and Reis [400]. The solution containing the peptides was collected and stored at -20°C until application to a MALDI plate. Peptides were acidified with trifluoroacetic acid (TFA) and concentrated using C18 micro-columns (C18 Tips, 10 µl, Thermo scientific, 87782). Peptides were thereafter eluted from the micro-column directly onto the MALDI plate with 1.5 µl of α -CHCA matrix (8 mg/mL) prepared in acetonitrile (50%, v/v), TFA (0.1%, v/v) and 6 mM ammonium phosphate. MALDI mass spectra were externally calibrated following the manufacturer's instructions (TOF/TOF calibration mixture, AB SCIEX) and internal calibration was applied using trypsin autolysis peaks. Peptide mass spectra data was collected in positive ion reflector mode in the range of m/z 700-4000 (4800 Plus MALDI TOF/TOF Analyzer, AB SCIEX).

Proteins were identified by combining Peptide Mass Fingerprint and MS/MS information. Proteins were searched in a locally stored NCBI copy of protein sequences of the genomes of the sea anemones *Exaiptasia pallida* (26,042 protein count, GenBank accession: GCA_001417965.1) and *Nematostella vectensis* (24,780 protein count, GenBank accession: GCA_000209225.1), using the Mascot search engine (Version 2.4). The search included peaks with a signal-to-noise ratio greater than 10 and allowed for up to two missed trypsin cleavage sites, mass tolerance of 50 ppm, cysteine carbamidomethylation (fixed modification), methionine oxidation (variable modification), and a charge state of +1. For a match to be considered significant, protein scores with a probability greater than 95% ($p < 0.05$), calculated by the Mascot software, were required [400]. The data generated from 2D-MALDI procedures were also searched against UniProtKB protein sequence database in the Metazoa section [401, 402], using the same parameters mentioned before.

In solution protein digestion and MS/MS analysis

For LC-MS/MS analysis, SF and IF protein samples were processed by filter aided sample preparation (FASP) method [403] with the following modifications. Protein samples (40 µg) were alkylated and digested with trypsin (recombinant, proteomics grade, Roche), at enzyme to protein ratio of 1:100 (w/w), for 16 h at 37°C, in centrifugal filter units with nominal molecular weight limit (NMWL) of 30 kDa (MRCF0R030, Millipore, Billerica, MA, USA). Peptides were subsequently recovered by centrifugal filtration, acidified with formic acid (FA) (10%, v/v), desalted and concentrated by reversed-phase extraction (C18 Tips, 100 µl, Thermo scientific, 87784) using acetonitrile (ACN) (70%, v/v) and TFA (0.1%, v/v) for peptide elution. Before LC-MS/MS, the peptides were recovered in 0.1% (v/v) Formic acid (FA) to the concentration of 0.04–0.06 µg/µl.

FASP protein digests (duplicate samples) were analyzed by nano-LC coupled to a hybrid Ion-trap mass spectrometer (LTQ Orbitrap Velos Pro –ETD, Thermo Scientific,

Waltham, MA, USA) as described previously [229]. Peptides were separated by reverse-phase chromatography (20 mm × 100 µm C18 precolumn followed by a 100mm× 75 µm C18 column with particle size 5 µm, NanoSeparations, Nieuwkoop, Netherlands) using a linear ascending gradient of buffer B (ACN + FA, 0.1%, v/v), being buffer A TFA, 0.1%, v/v in water. The gradient started from 2% B to 30% B in 40 min and to 95% B (v/v) in 30 min, at a flow rate of 0.3 µL/min (total elution time 70 min). Peptides were analyzed by on-line nano-electrospray ionization (easy nano-ESI) in positive mode, with Xcalibur software (version 2.6, Thermo Scientific). Full scans were performed at a resolution of 30 000 with scan ranges of m/z 380–2000. The top 20 most intense ions were isolated and fragmented with CID by applying normalized collision energy of 30% value, isolation width of 2.0, activation time of 10 milliseconds and Q-value of 0.25. In total 4 nano-LC-MS/MS runs were performed.

Peptide identification

LTQ raw data were searched against custom cnidarians protein databases using MaxQuant freeware software (version 1.5.5.1) with the Andromeda search engine. MS and MS/MS tolerances were set to 10 ppm and 0.6 Da, respectively. Trypsin was selected for protein cleavage allowing for one missed cleavage. Carbamidomethylation and oxidation were selected as static and dynamic modifications, respectively. Identifications were validated by performing a decoy database search for the estimation of False Discovery Rate (FDR) and peptide identifications were accepted if they could be established at a probability greater than 95.0%. Protein identifications were accepted if they could be established at a probability greater than 99.9% and contained at least two identified peptides (Razor + unique peptides) [201, 202], based on Occam's razor principle). The protein database utilized was the locally stored NCBI copy of protein sequences of the genomes of the sea anemones *E. pallida* (26,042 protein count, GenBank accession: GCA_001417965.1), *N. vectensis* (24780 protein count, GenBank accession: GCA_000209225.1), *Hydra vulgaris* (21,993 protein count, GenBank accession: GCF_000004095.1) and *Acropora digitifera* (33,878 protein count, GenBank accession: GCF_000222465.1). The identification of potential toxins was done against the manually reviewed venom proteins and toxins database, from the animal toxin annotation project of the UniProtKB/Swiss-Prot protein knowledgebase [205, 404, 405], available at <http://www.uniprot.org/program/Toxins> (database size 1.20 MB, downloaded on June 16, 2016).

Protein homology search and GO analysis

Protein sequences with unknown function were annotated with a blast search in the National Centre for Biotechnology Information database (NCBI, <http://www.ncbi.nlm.nih.gov/>) using blastp algorithm employing a threshold e-value of 1×10^{-10} . Total of proteins identified with Maxquant software, were also blasted and mapped using the Blast2Go software

(version 2.4.4, <http://www.blast2go.com/>) [228]. Gene ontology (GO) terms were used to group proteins within the domains of biological process (BP), cellular component (CC), and molecular function (MF).

4.4 Results and Discussion

2DE and MALDI-TOF/TOF analyses

The gel-based proteome analysis revealed 61 and 36 spots from the SF and IF, respectively. From the spots analyzed by MALDI-TOF/TOF, 23 peptide sequences belonging to eight proteins were identified in the SF, approximately 38% of the total analyzed (Figure 4.1, Table 4.1). Proteins identified in the SF comprised five different enzymes: Superoxide dismutase, Triosephosphate isomerase, Ribonuclease, two Fructose-bisphosphate aldolases and Alpha-enolase. In addition, Peroxiredoxin and two Ferritins were identified. However, three of these proteins matched to “predicted protein” as best hit, but were then further annotated using blastp algorithm in the NCBI with the accession number retrieved from the custom sea anemones databases. Unlike shotgun proteomics, for gel-based analysis were used only two sea anemones databases, since additional search was carried against UniProtKB/Swiss-Prot in the Metazoa section. However, best results corresponded to local analysis. On the other hand, no proteins were identified with statistic confidence from the IF (Supplementary Figure S4.1) and in both cases SF and IF, the use of different database like UniProtKB/Swiss-Prot did not improved the identification. The details of blast search and protein identification by MALDI-TOF/TOF mass spectrometry of the protein identified from the 2DE is shown in Table 4.1. It is noteworthy, that some of the proteins identified have been previously reported in other cnidarians [155, 406, 407], but constitute the first report for *B. verrucosa*.

The identification rates obtained for SF are like those reported in previous studies of other marine species, when comparable proteomics protocols were used [293, 398, 408]. On the other hand, the absence of identifications in IF was unexpected, and this is evidence that our proteomics protocol is likely not optimized for the analysis of the type of proteins present in this fraction. Since IF may be enriched with hydrophobic membrane proteins, the lack of identifications may be related, among other possible causes, to the inefficient digestion of these proteins with trypsin, thus hindering the generation of proteolytic peptide fragments for MS/MS sequencing analysis. This limitation of trypsin when cleaving these proteins particularly in the hydrophobic and transmembrane domains can be overcome by combining the activities of other proteases [409, 410].

The identified proteins seem to play important roles related with RNA degradation, glycolysis and antioxidant pathways. Moreover, some proteins like alpha aldolase seem to

play diverse molecular and physiological roles. In fact, several antibacterial, antiparasitic, antifungal and autoantigen activities have been proposed [411]. Alpha aldolase expression and activity have been associated with the occurrence and metastasis of cancer, as well as with growth, development and reproduction of organisms [411]. Its expression seems to be related to heat shock [412], but it is also probably active under anaerobic condition [411]. In general, some of these proteins act as stress protein against environmental changes by exerting a protective effect on cells.

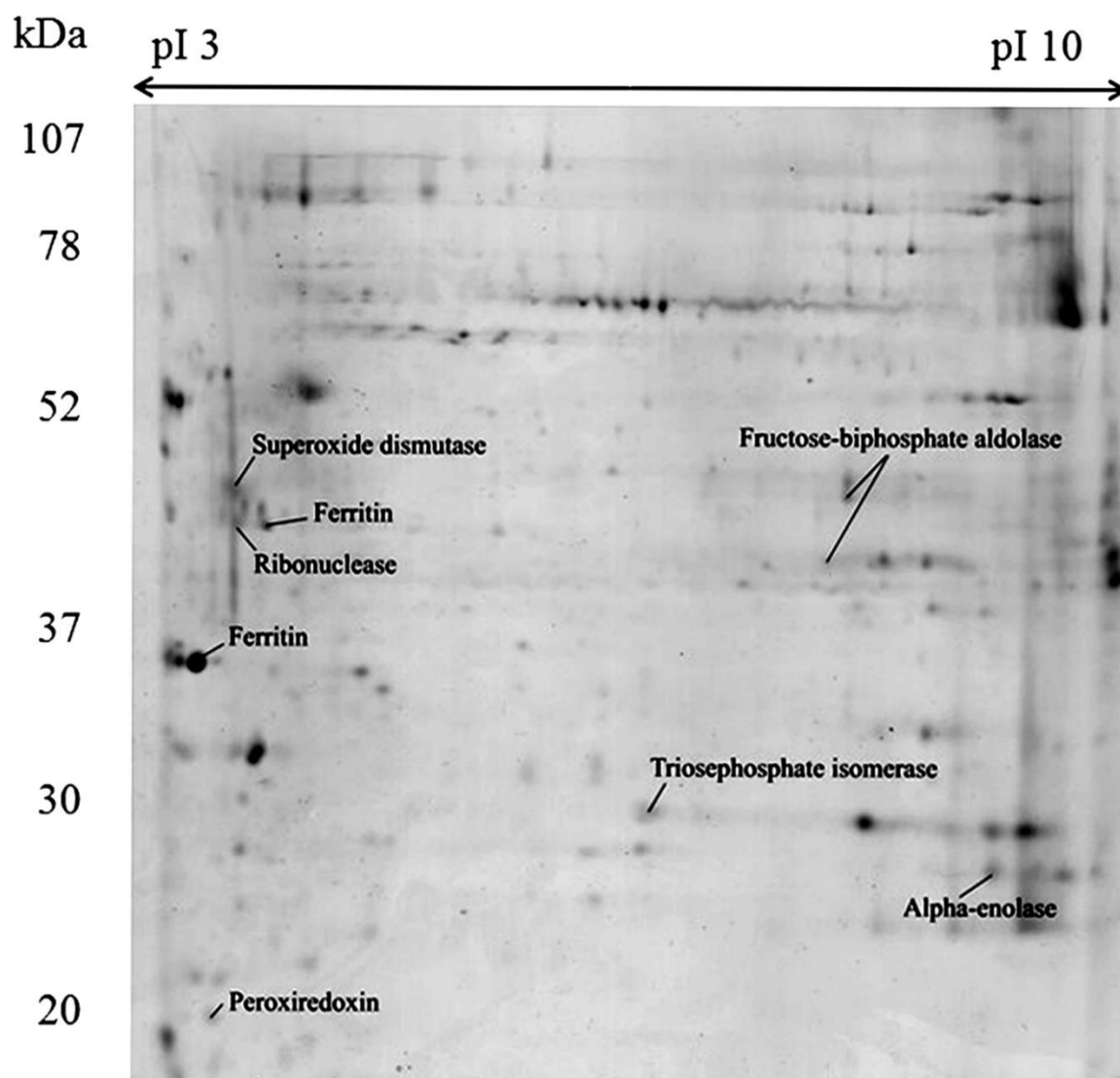


Figure 4.1. Two-dimensional gel electrophoresis and identification of soluble proteins from *Bunodactis verrucosa*. The first dimension was carried out on 17 cm, pH 3-10 IEF gel strips and the second dimension on 12% SDS-PAGE gels. Gels were stained with colloidal Coomassie blue G-250. Identified proteins are indicated with the most commonly used name.

Ribonucleases, also known as RNases, are common and widely distributed catalytic proteins among animals, involved in the RNA degradation [413]. Three different RNases were detected: Triosephosphate isomerase, Fructose-bisphosphate aldolase and Alpha-enolase, which are involved in the glycolytic pathway. Triosephosphate isomerase is a glycolytic enzyme that catalyzes the interconversion of the three-carbon sugars such as dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate [414]. Aldolases are stereochemistry-specific enzymes acting in a diverse variety of condensation and cleavage reactions [411]. Specifically, Fructose-1,6-bisphosphate aldolase is involved in gluconeogenesis and glycolysis, controlling the production of fructose-1,6-bisphosphate from the condensation of dihydroxyacetone phosphate with glyceraldehyde-3-phosphate [415, 416]; while Alpha-enolase is a versatile metalloenzyme that catalyzes the conversion of 2-phosphoglyceric acid to phosphoenolpyruvic acid [411].

Table 4.1: Blast Search summary. The table shows a summary of the data concerning protein identification by MALDI-TOF/TOF mass spectrometry of the proteins separated in two-dimensional gel electrophoresis.

protein name ¹	Species ²	Protein ³ score	Accession ⁴ number	Ion ⁵ score	Peptide sequence ⁶
predicted protein (Peroxiredoxin)	<i>Nematostella vectensis</i>	137	XP_001640260.1	15	R.LIQAFQFTDK.H
				115	K.DYGVLLDQGVLR.G
Ferritin	<i>Nematostella vectensis</i>	124	XP_001632011.1	114	R.QNYHEECEAGINK.Q
Ferritin	<i>Nematostella vectensis</i>	117	XP_001627474.1	11	K.LMKFQNQR.G
				97	R.QNYHEECEAGINK.Q
predicted protein (Ribonuclease)	<i>Nematostella vectensis</i>	106	XP_001634183.1	93	R.VEIEAIAIVGEVKDE.-
Superoxide dismutase [Mn]	<i>Exaiptasia pallida</i>	428	KXJ18609.1	76	K.DFGSFENFK.X
				67	K.KDFGSFENFK.
				103	K.AYDVIDWTVNADR.Y
Triosephosphate isomerase	<i>Nematostella vectensis</i>	356	XP_001633516.1	56	K.FFVGGNWK.M
				22	R.KFFVGGNWK.M
				95	K.VIACIGELLSER.E
				19	R.NIFGEKDELIGEK.V
				121	K.VVIAYEPVWAIGTK.T
predicted protein/Alpha-enolase	<i>Nematostella vectensis</i>	95	XP_001632906.1	10	K.YNQLLR.I
				37	R.AAVPSGASTGIYEALRLR.D
				10	K.LAMQEFMLLPTGASNFR.E
Fructose-bisphosphate aldolase	<i>Nematostella vectensis</i>	151	XP_001629735.1	41	K.LTFSFGR.A
				23	R.LLRDQGIIPGIK.V
				28	R.LANIGVENTEENRR.L
				24	R.LLRDQGIIPGIKVDK.G
Fructose-bisphosphate aldolase	<i>Nematostella vectensis</i>	97	XP_001629735.1	28	K.LTFSFGR.A
				32	R.LANIGVENTEENRR.L

¹best hit NCBI accession number; ²name of the species best hit belongs; ³Score obtained for the MS ion; ⁴ NCBI accession number retrieved from the custom database; ⁵ MASCOT's score for ion peptides; ⁶ peptides sequences identified with statistical significance

On the other hand, Ferritin is one of the most important proteins in iron metabolism, acting as primary iron storage protein or iron transporter, solubilizing iron and thus regulating its homeostasis [417, 418]. Peroxiredoxin, also called thioredoxin peroxidase or alkyl hydroperoxide reductase, has been proposed as antioxidant protein [419-421]. Both proteins seem to play an important role by protecting the cells against reactive oxygen species [422], so they are likely to be natural anti-UV radiation agents [423]. Similarly, Superoxide dismutase is another relevant antioxidant protein [422, 424]. The high expression of this protein as part of the antioxidant defense system makes sense, since aerobic organisms need to deal with oxygen species produced as a consequence of aerobic respiration and substrate oxidation [424].

Protein identification from shotgun proteomics analyses

A methodology based on shotgun analysis was employed to investigate the whole-body proteome of *B. verrucosa*. This methodology has been previously reported as suitable for diverse purposes related to protein identification such as characterization of complex sample, inference of the main enzymatic pathway involved in a tissue, even to reveal venom composition [229, 230, 425, 426]. Altogether, 688 peptide sequences were identified between the two replicates of the fractions analyzed (SF and IF), which accounted for 412 groups of non-redundant proteins (Supplementary Table S4.1) retrieved from custom cnidarians databases. Of all protein detected, 97 were identified from two or more peptides. Only four proteins were detected as potential contaminants in the first search against custom database, while 69 sequences accounted for 35 putative proteins as contaminants against UniProtKB/Swiss-Prot database (Supplementary Table S4.2). Of such contaminants, 10 proteins were identified from two or more peptides and were related mostly to human keratin and trypsin. In the case of contaminants, proteolytic fragments from trypsin and keratin were the most commonly found, which are difficult to avoid and thus are ubiquitous in proteomic analysis [427]. The functional annotation of all proteins (except for contaminants) was further addressed.

The fact that several IF proteins were identified by this shotgun method shows the increased potential of this method over 2DE/MALDI-TOF/TOF for the analysis of membrane proteins, even when carried out on the basis of the activity of a single protease (trypsin).

All proteins identified from the gel-based analysis were also found among those identified by the shotgun proteomic analysis. As an example, the shotgun analysis allowed the identification of Peroxiredoxin (XP_001640260.1, see Table 4.1) from two peptides sequences belonging to different organisms (Supplementary Table S4.3): one peptide matched Peroxiredoxin-4 (KXJ19217.1) from *E. pallida*, and the second one Peroxiredoxin-4 (KXJ22794.1) from *E. pallida* and peroxiredoxin-like isoform X2 (XP_015769163.1) from *A.*

digitifera. In the case of Peroxiredoxin-4 (KXJ22794.1), four peptides were identified for the protein and four for the protein groups (razor + unique, see file Additional_shotgun_terms in Supplementary Materials/Additional files). However, only nine peptides generated by MALDI-TOF/TOF fragmentation from gel spots, were also detected within peptides resulting from the Orbitrap's approach. Despite the less number of protein identified from 2DE gel, this methodology represented a complement for shotgun proteomics analyses, increasing the number of peptides for the reconstruction of each protein. In fact, in 2D-MALDI fingerprint approach the number of peptides matching some proteins such as Superoxide dismutase (KXJ18609.1), Alpha-enolase (XP_001632906.1) Triosephosphate isomerase (XP_001633516.1) and both Fructose-bisphosphate aldolase (XP_001629735.1; XP_001629735.1), were higher than number of peptides per such protein in case of LC-MSMS.

Gene Ontology (GO) Annotation

Of the 412 proteins identified with Maxquant software, 408 were successfully mapped using the Blast2Go software. The remaining four proteins without analysis corresponded to potential contaminants. Out of the total of proteins analyzed, 149 proteins were successfully annotated, representing the 36.5 % of the total. Thus, 259 proteins remained without GO annotation, of which only four proteins were blasted without hits, 36 were mapped and 219 comprised positive hits. In total, 223 proteins were not included into the GO annotation considering the level 2 of protein classification, likely due to the absent of similar protein sequences in the protein databases. Moreover, most of these proteins retrieved as hit from cnidarian databases were "predicted". This result confirms that limited information its known about sea anemones and cnidarians products.

Among the four databases analyzed, most hits corresponded to the species *E. pallida*, followed by *N. vectensis* (Figure 4.3), as expected according to its relative phylogenetic position [428], although *E. pallida* has the largest number of proteins among the databases used. Afterwards, the proteins identified as positive hits were functionally annotated per the Gene Ontology (GO) (<http://www.geneontology.org/>) nomenclature. Then, GO terms were assigned to each contig and annotated per GO Distribution by Level (2), regarding the three major GO categories: BP, MF and CC.

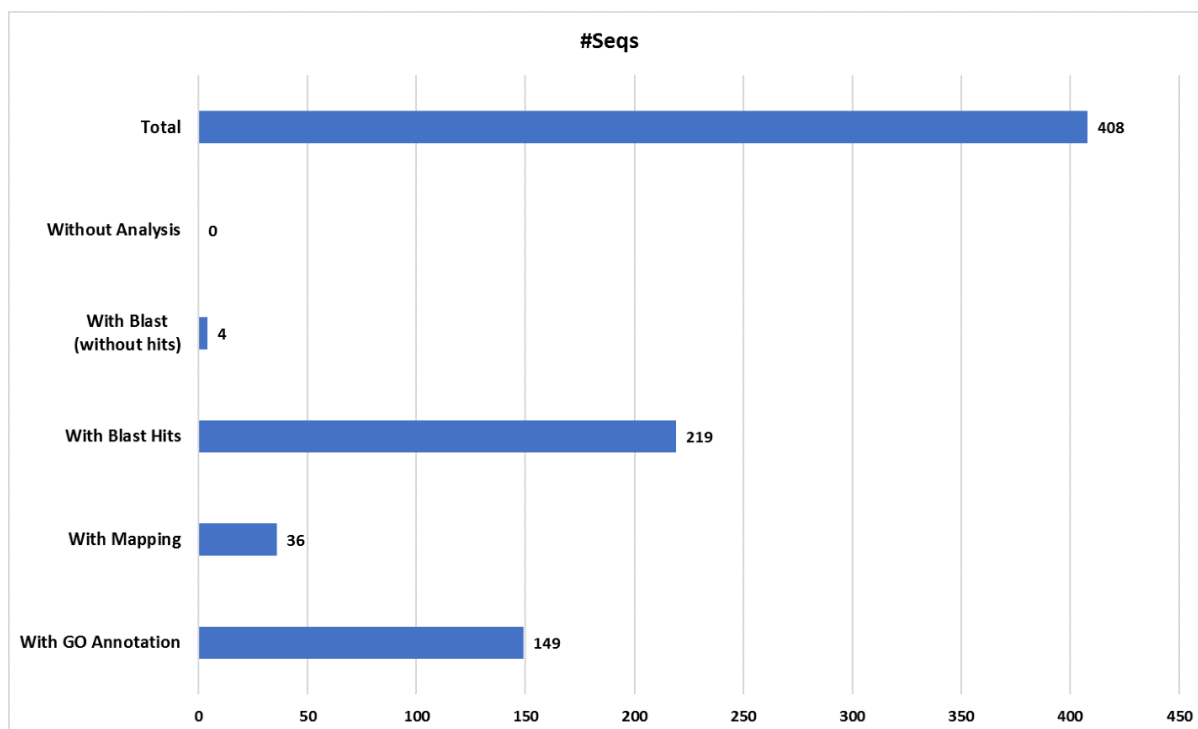


Figure 4.2. Blast2Go Data Distribution chart. Included in the figure is the statistics of annotation performed with Blast2Go software against the four cnidarian databases analyzed

The groups of proteins obtained from high-throughput analyses were classified per Blast2Go software, considering the GO Distribution by Level (2) (Figure 4.4). The most represented of GO terms in the category of Biological Process (BP) was metabolic process (GO:0008152), followed by cellular process (GO:0009987) and single-organism process (GO:0044699). In the case of Molecular Function (MF), the most matched GO terms were binding (GO:0005488), catalytic activity (GO:0003824) and structural molecule activity (GO:0005198), in this order; while in the category of Cell Component (CC) the most significant were cell part (GO:0044464), cell (GO:0005623) and organelle (GO:0043226). It is noteworthy that some proteins can be included in more than one GO term, since each protein could play diverse roles. Details of GO annotation and protein accession number can be found in the Supplementary Figure S4.2 and Supplementary Table S4.5.

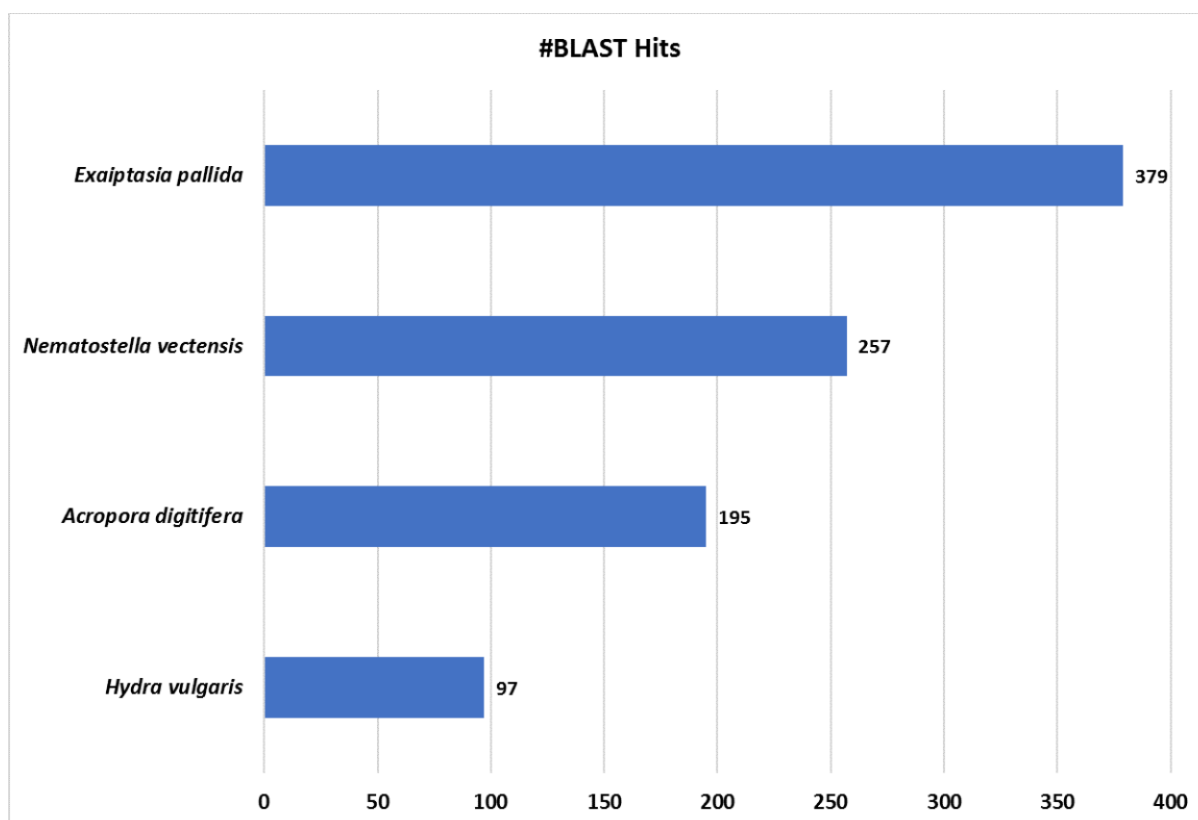


Figure 4.3. Blast2Go Species distribution chart. In this figure, it is shown the number of blast hits retrieved from the four cnidarian databases analyzed.

Among the 111 proteins matched to the GO term BP, 85 proteins (76.56%) classified as metabolic process, 77 (69.37%) for cellular process and 44 (39.64%) as single-organism process. In this group, in the GO level 3, 64 proteins were related with the GO name of “primary metabolic process” and “organic substance metabolic process”, both belonging to “metabolic process” as parent. Besides, 52 proteins were associated with “cellular metabolic process”, which were involved in both metabolic process and cellular process as parents (for details of GO annotation see Supplementary Figure S4.2, Supplementary Table S4.5).

In total, 86 proteins were included in the category of the CC. Among them, 76 proteins (88.37) matched for “cell part” and “cell”. However, this is an ambiguity, since all sequences detected as “cell part” are part of the “cell” (Supplementary Figure S4.2). Although, other proteins represented by the sublevels, related to cytoplasmic elements as part of intracellular components, were also subcategories of the “cell”. The GO “intracellular” was more represented with 73 proteins (84.88%) in level 3 than those “organelle” and “membrane” in the superior level 2, with 51 proteins (59.3%) and 33 proteins (38.37%), respectively.

In addition, 135 proteins were grouped in the MF category. Among them, “binding” with 99 proteins (73.3%) was the most significant. In this group, a total of 66 (48.89%)

proteins were involved in “ion binding”, whereas both “heterocyclic compound binding” and “organic cyclic compound binding” hit 62 proteins (45.93%). The second most significant GO term “catalytic activity” comprised 71 proteins (52.59%), of which the most remarkable function was “hydrolase activity”, accounting for 37 proteins (27,41%) acting mainly on acid anhydrides, in phosphorus-containing anhydrides. Moreover, 18 of these enzymes were involved in pyrophosphatase activity, of which 17 were associated with nucleoside-triphosphatase activity (Supplementary Figure S4.2).

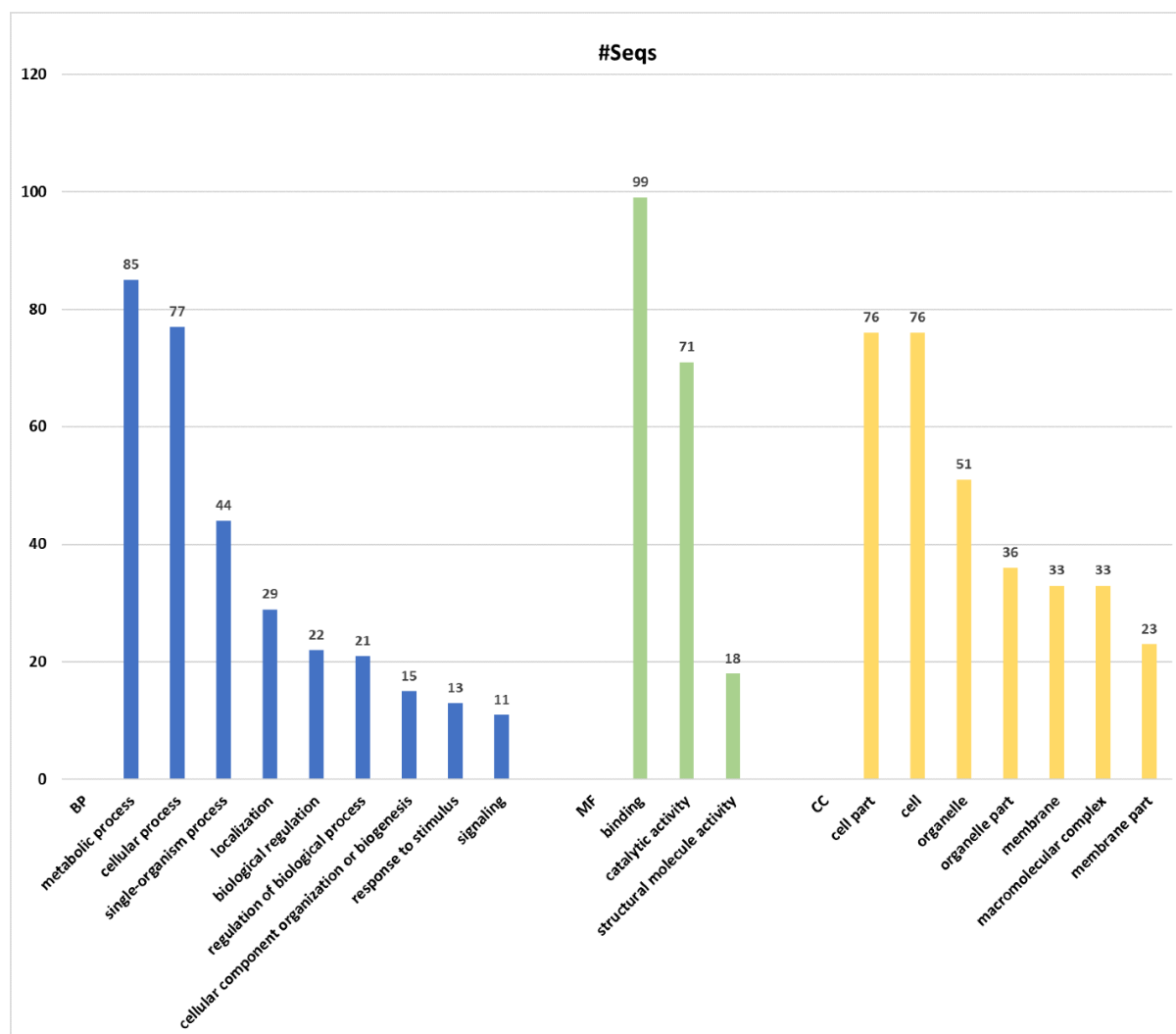


Figure 4.4. Blast2Go hits Gene Ontology (GO) annotation. The Figure shows the blast hits annotation considering the three major GO Categories of GO Distribution by Level (2): Biological Process (PB) in blue, Molecular Function (MF) in green and Cellular Components (CC) in yellow.

Top KEGG pathways

On the other hand, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses revealed 28 enzymes involved in 41 different pathways. The accession number of the protein involved in each pathway and other details can be found in Supplementary Table

S4.6. Considering the number of protein matched, the most relevant pathways were Purine and Thiamine metabolism, with 18 and 17 proteins matched, respectively (Table 4.2). In addition, three enzymes: adenylypyrophosphatase, phosphatase and RNA polymerase were found to be involved in the Purine metabolism pathway, whereas only a phosphatase resulted in the Thiamine metabolism. The Purine metabolism pathway is close related to the metabolism of nucleotide [429], since purine constitutes subunits of nucleic acids and precursors for the synthesis of nucleotide cofactors, while Thiamine metabolism pathway is fundamental in the metabolism of carbohydrates [430].

Table 4.2. Top twenty KEGG pathways.

Pathway	#Proteins in the pathway	#Enzymes in Pathway
Purine metabolism	18	3
Thiamine metabolism	17	1
Biosynthesis of antibiotics	14	13
Glycolysis / Gluconeogenesis	9	6
Carbon fixation in photosynthetic organisms	9	6
Amino sugar and nucleotide sugar metabolism	6	3
Methane metabolism	6	3
Pyruvate metabolism	5	4
Cysteine and methionine metabolism	4	5
Citrate cycle (TCA cycle)	4	3
Fructose and mannose metabolism	4	2
Various types of N-glycan biosynthesis	4	1
Glycosphingolipid biosynthesis - ganglio series	4	1
Glycosaminoglycan degradation	4	1
Glycosphingolipid biosynthesis - globo and isoglobo series	4	1
Other glycan degradation	4	1
Glyoxylate and dicarboxylate metabolism	3	3
Carbon fixation pathways in prokaryotes	3	2
Pentose phosphate pathway	3	2
Histidine metabolism	2	2

Interestingly, one of the most significant among the top twenty pathways was the Biosynthesis of antibiotics. In the latter pathway, a total of 14 proteins, accounted for 13 enzymes grouped in five major families: dehydrogenase, transaminase, carboxykinase(GTP), hydratase, isomerase and aldolase. Most of the protein matched in this pathway belongs to larval stage of *N. vectensis*. This result is of particular interest, because of the abundance of proteins involved in defenses against pathogens, during the most vulnerable stage in the life cycle. Thus, this finding supports that sea anemones may be considered as a promising source of antibiotic compounds [431-433]. Other relevant pathways were Glycolysis/Gluconeogenesis and Carbon fixation in photosynthetic

organisms, both involved in the production of energy. The presence of proteins associated with Carbon fixation in photosynthetic organisms is likely due to symbionts such as zooxanthellae, which are known to be present in sea anemones [434, 435].

The isomerase detected in the Biosynthesis of antibiotics pathway, was the same to that identified in the gel-based analyses as Triosephosphate isomerase from *N. vectensis* (XP_001633516.1). This is also involved in other pathways like Glycolysis/Gluconeogenesis, Carbon fixation in photosynthetic organisms, Fructose and mannose metabolism and Inositol phosphate metabolism. The predicted protein (XP_001632906.1), homologue to Alpha-enolase, and the Fructose-bisphosphate aldolase (XP_001629735.1) from *N. vectensis* were both involved in the pathways of Biosynthesis of antibiotics and Glycolysis/Gluconeogenesis. In addition, the mentioned predicted protein was also found in the Methane metabolism pathway, while the Fructose-bisphosphate aldolase also occurred in some pathways like Carbon fixation in photosynthetic organisms, Methane metabolism, Pentose phosphate pathway and Fructose and mannose metabolism. In general, these analyses support the diverse roles of some of the proteins identified, given additional information related to its biological function.

Detection of potential toxins

Among all peptides detected, 63 sequences matched for 58 potential toxins (Supplementary Table S4.2, Supplementary Table S4.4), but only five toxins with more than one peptide (Table 4.3). Specifically, the five proteins matched as potential toxins were retrieved from different species other than cnidarians and each reconstructed by two peptide sequences. Besides, these peptides were not redundant to those proteins reconstructed from the previous analyses with the four cnidarians database. In fact, the origin of such peptides by fragmentation of the protein matched as potential toxin (Table 4.3), which represents a better explanation for our results. Therefore, it is unlikely a false-positive assumption that the peptides were generated from proteins related to potential toxins.

The proteins identified as potential toxins comprises several previously reported toxins and other non-reported in cnidarians. Herein, we found two proteins related to metalloproteinases, one zinc metalloproteinase/disintegrin (VM2M2_DEIAC) of the snake *Deinagkistrodon acutus* [436] and other called Neprilysin-1 (NEP_TRILK) from brush-footed trapdoor spider *Trittame loki* [437]. Both proteins represent two of the three classes of metalloproteinases found in Hydra genome: astacin class, matrix metalloproteinase class, and neprilysin [438]. Metalloproteinases have been subsequently reported in hydra [439, 440], jellyfish [22, 161, 441], but less in sea anemones [217]. Its structure and function seem relatively conserved among metazoans [441], since it can play a broad range of roles in biological process related to hydrolytic functions and development [438]. However, the

peptides obtained matched specifically to proteins, which have been proposed as venom components [436, 437]. In general, the most significant role of the protein found in the present work, must be related to its capacity of breakdown the extracellular matrix [438]. Moreover, this protein holds gelatinolytic and fibrinolytic activities, as previous reported from the venoms of four Scyphozoan jellyfish [161].

Table 4.3. Potential toxins. The table shows potential toxins identified by MaxQuant software against the venom section of UniProtKB/Swiss-Prot database.

Protein ¹ name	Species ²	Score ³	Accession ⁴ number	Ion ⁵ score	Peptide sequence ⁶	Fraction ⁷ (Rep.)
SE-cephalotoxin	<i>Sepia esculenta</i>	11.47	CTX_SEPES	62.7	AGYIMGNR	IF (1)
				42.8	LDQINDKLDK	IF (1)
Basic phospholipase A2 vurtoxin	<i>Vipera renardi</i>	12.06	PA2B_VIPRE	2.9	CCFVHDCCYGNLPDCNPKIDR	SF (1)
				18.3	NGAIVCGK	IF (1)
Alpha-latroinsectotoxin-Lt1a	<i>Latrodectus tredecimguttatus</i>	11.73	LITA_LATTR	22.7	EMGRKLDK	IF (2)
				3.01	NSCMHNDKGCCFPWSCVCWSQTVSR	SF (1)
Zinc metalloproteinase/disintegrin	<i>Deinagkistrodon acutus</i>	11.48	VM2M2_DEIAC	27.4	FPYQGSSIILESGNVNDYEVVYPRK	SF (1)
				31.7	NTLESFGEWRAR	IF (1)
Nepirylsin-1	<i>Tritame loki</i>	11.49	NEP_TRILK	28.4	LAHETNPR	IF (1)
				71.3	LEAMINK	SF (2)

¹UniProtKB/Swiss-Prot name of the protein identified as potential toxin; ²name of the species best hit belongs; ³Protein score which is derived from peptide posterior error probabilities; ⁴UniProtKB/Swiss-Prot hit accession number; ⁵Andromeda score for the best associated MS/MS spectrum; ⁶UniProtKB/Swiss-Prot accession number; ⁷fraction (IF: Insoluble fraction; SF: Soluble fraction) where a peptide was detected and replicates they occurred.

Another enzyme detected was a phospholipase A2 (PLA₂) called vurtoxin (PA2B_VIPRE) from the steppe viper *Vipera renardi* [442]. Phospholipases A2 are commonly found in the venom of the most toxic animals like cnidarians, cephalopods, insects, arachnids, and reptiles [2]. Cnidarians PLA₂ shows a significant phylogenetic distance to higher metazoans PLA₂s, and have been proposed as the ancestors [91]. In general, PLA₂s can act in the arachidonic pathway or in the calcium-dependent hydrolysis of the 2-acyl groups in 3-sn-phosphoglycerides, showing a preference for phosphatidylglycerol over phosphatidylcholine [121, 443]. Despite structural difference, its biological role prevails in cnidarians, but its role in reptiles has been revealed most as antiplatelet, myotoxic, and neurotoxic activities [121, 444]. Specifically, vurtoxin showed homology with the neurotoxic PLA₂ ammodytoxins [442]. However, it is not clear if this toxin can act as neurotoxin in this species, since vurtoxin occurred as a minor component in the venom of *V. renardi* [442].

In addition, two putative neurotoxins named as Alpha-latroinsectotoxin-Lt1a (LITA_LATTR) and SE-cephalotoxin (CTX_SEPES) were identified. The first one mentioned, also known as alpha-LIT, was purified from venom glands of the Mediterranean black widow spider *Latrodectus mactans tredecimguttatus* [441]. The proposed mechanism of toxicity

involved presynaptic effects, acting selectively only for insects [441]. On the other hand, SE-cephalotoxin have been characterized from the salivary gland of cuttlefish *Sepia esculenta* [161]. The lethality of this toxin was very high to crab, seemingly by neurotoxic mechanism, since the symptoms caused loss of movement, flaccid paralysis and even death [161]. However, SE-cephalotoxin has been considered as a new class of proteinaceous toxin, due to the lack of homology with any other toxins, even those cephalotoxins from octopuses [161]. Therefore, the evidences of a potential SE-cephalotoxin from *B. verrucosa*, constitutes a highlighted finding as the second report of this toxin.

Furhermore, others 53 non-redundant peptide sequences matched to 53 potential toxins (Supplementary Table S4.2). Of all, 21 peptides sequences matched to 21 potential neurotoxins comprising presynaptic and postsynaptic toxins like ion channel blockers, mostly voltage-dependent potassium and calcium channels. Among them was found a Kunitz-type serine protease inhibitor, which can act as inhibitor of both serine proteases and voltage-gated potassium channels (Kv) [217]. Besides, three metalloproteases, two hyaluronidases, and a Beta-fibrinogenase were detected. On the other hand, seven PLA₂s and three PLA_D occurred within potential toxins. Another potential toxin identified with PLA₂s activity, was the Helofensin-1 characterized from the genus *Heloderma* [445, 446]. This toxin has no hemorrhagic nor hemolytic activities, instead directly inhibited the electrical stimulation of the isolated hemi-diaphragm of mice [445]. Finally, four hemolitics/cytolytic proteins and five additional proteins involved in the coagulation pathway (including two “snaclec”) were found.

Putative use of toxins by *B. verrucosa* in prey catching and feeding

Sea anemones are ancient active predators, belonging to what is considered “the oldest extant lineage of venomous animals” [447]. The *B. verrucosa* inhabits tidepools in rocks, crevices in shallow water [448], where occurs mussels, small crabs, and goby fishes as potential preys. This sea anemone feeds at least on mussels, since we found specimens regurgitating one or more empty mussel shells, after removal from the substrate during sampling. In the sampling area mussels were abundant covering rocks, even in the pools where sea anemones grow. Therefore, these bivalves may constitute the main food source for *B. verrucosa*. This is not an isolated fact, since mussels seem to be the main food source for others intertidal sea anemones like *Anthopleura elegantissima* and *Anthopleura xanthogrammica* [449, 450]. Moreover, mussels are suitable to be fed by sea anemones in home aquariums [451]. However, bivalves can close their valves for prolonged periods of time under adverse environmental condition [452, 453]. In other words, how can sea anemones obtain nourishments from mussels, if these bivalves tightly close the valves when feel the predator attack?

Mussels are abundant in the intertidal community and their movements are limited. In this scenario, sea anemones can capture a close mussel with its tentacles and introduce it into the gastrovascular cavity. Once the mussel is captured, it immediately closes its valves and stops filtering. Nonetheless, the sea anemones have cnidocytes in the gastrovascular cavity [156] capable of break mussels' protection. First, hydrolytic enzymes like zinc metalloproteinase/disintegrin, hyaluronidases and proteases found in *B. verrucosa* may be poured into the gastrovascular cavity. The combination of such enzymes could degrade the tissues that seals the shell, probably a dorsal elastic proteinaceous-ligament extending for the length of the hinge [454]; or through the ventral margin of the mussel. The tissues degradation by metalloproteinases can facilitate the diffusion of neurotoxins inside the prey. Then, neurotoxins could act on the adductor muscle, which the loss of function will lead to valve opening.

Specifically, SE-cephalotoxin can penetrate inside the valves, inhibiting the adductor muscles, thus producing flaccid paralysis increasing the valves gape aperture. The high solubility previously reported for SE-cephalotoxin seems to play an important role in the diffusion of this toxin in sea water. This property should be useful whether preys are nearby the sea anemone, because SE-cephalotoxin could disperse around or in the sea water remnant inside the shell after enclosed its valves. Besides, this feature can be used as an advantage to subdue prey prior to eating. Other neurotoxins detected, and the PLA₂ vurtxin, are also able to block the adductor muscles. However, the diversity of toxins found is likely related to others potential preys as crabs and goby fishes (Gobiidae, Perciformes), polychaetes worms and starfish. Interestingly, other cephalotoxins have been previously purified from species of octopodiform cephalopods [455-458]; which are likely used to neutralize crabs and bivalves. Altogether, toxins found seemingly act synergistically to subdue mussels. Indeed, a similar mechanism in which hydrolytic enzymes like metalloproteinase facilitate the access of neurotoxic peptides to synaptic targets was previously proposed in the spider *T. loki* [437].

4.5 Conclusion

This work revealed the first draft of the whole proteome of the sea anemone *B. verrucosa*. The shotgun proteomics analysis yielded most of the protein identified with a total of 412. Altogether, both gel-based and gel-free approaches of proteomics analyses and functional bioinformatics analyses revealed three major groups of proteins belonging to "metabolic process", "binding" and "cell parts" GO categories. Unlike throughput analyses, only eight proteins were identified from two-dimensional electrophoresis combined with MALDI-TOF/TOF. These eight proteins comprised enzymes mainly involved in the glycolytic pathway, antioxidants activities and RNA degradation. Notably, according to the

results of KEEG analysis a significant number of enzymes corresponded to the Biosynthesis of antibiotics pathway indicating the importance of the biological antimicrobial chemical defense mechanisms. Moreover, we identified potential toxins: metalloproteinases, and neurotoxin like SE-cephalotoxin. The combination of proteomic evidences and the ecology of the species, shed light about its strategy to subdue preys like mussels. In this sense, the toxins seemingly act synergically. Metalloproteinase may produce a degradation of the tissues, aiding the diffusion of the neurotoxins to the target, producing muscle paralysis. Hence, this work constitutes a reference proteome for future studies in sea anemones, also given insight about its potential toxin production and its putative mechanism of action in feeding.

4.6 Supplementary Materials

Supplementary Materials of the Chapter 4 are available online at the following link:
<https://goo.gl/WnANkU>

CHAPTER 5

Venomous repertoire involved in the
Harderian gland transcriptomes of
three snakes (Colubridae) from Cuba

Dany Domínguez-Pérez^{1,2†}; Jordi Durban^{3†}; Guillermin Agüero-Chapin^{1,2}; Javier Torres Lopez^{4,5}; Reinaldo Molina Ruiz⁶; Juan J. Calvete³; Vítor Vasconcelos^{1,2} and Agostinho Antunes^{1,2*}

¹ CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, Porto 4450-208, Portugal

¹ CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, Porto 4450-208, Portugal

² Biology Department, Faculty of Sciences, University of Porto, Rua do Campo Alegre, s/n, Porto 4169-007, Portugal

³ Instituto de Biomedicina de València, Jaume Roig, 11, 46010, València, Spain

⁴ Faculty of Biology, Havana University, 25 St 455, La Habana CP 10400, Cuba

⁵ Department of Ecology and Evolutionary Biology, The University of Kansas, 1345 Jayhawk Blvd., Lawrence, Kansas 66045, USA

⁶ Centro de Bioactivos Químicos, Universidad Central “Marta Abreu” de Las Villas, 54830, Santa Clara, Cuba

[†]authors contributed equally to this work

*To whom correspondence should be addressed: aantunes@ciimar.up.pt

Keywords: lipocalin, toxins, binding, vomerolfaction, defense; *Caraiba andreae*, *Cubophis cantherigerus*, *Tretanorhinus variabilis*, Dipsadinae, West Indies

This chapter will be submitted for publication

5.1 Abstract

The Harderian gland is a cephalic gland, widely distributed among vertebrates. In some snake species, this gland can be even larger than the eyes size. The Harderian gland is anatomically connected to the vomeronasal organ via the nasolacrimal duct. The function of this gland remains elusive, but it has been proposed to be mainly involved in the production of saliva, pheromones, thermoregulatory lipids and growth factors, among others. The “omics” approaches have proved to be a suitable tool to study tissue expression, thus allowing function inferences. Unlike snakes’ venom glands, which have been largely studied (because of their relevance as a source of new compounds for medical and drug discovery research), no proteomics or transcriptomics analyses have been done in the Harderian gland. Herein we profiled the Harderian gland’s transcriptome of three colubrids from Cuba: *Caraiba andreae*, *Cubophis cantherigerus* and *Tretanorhinus variabilis* using Illumina HiSeq2000 100 bp paired-end. Besides the ribosomal and cellular components, the most expressed transcripts found were related to transport/binding, lipocalin/lipocalin-like toxins, bactericidal/permeability-increasing protein-like and venom components, and snakes’ toxins. Indeed, several largely described snake-venom toxins classes were identified. The expression levels of some toxin-transcripts in the profiled Harderian glands were comparable to the snake’s toxins classes, previously found in snakes’ venom gland transcriptomes. This study constitutes the first transcriptome analysis of the Harderian gland in vertebrates, specifically in colubrid snakes. Our results revealed new insights about the Harderian gland’s function, suggesting a likely role in the transport of substances and defense mechanism; which is supported by the significant expression of binding/transport transcripts. Moreover, the detection of bactericidal proteins and toxin-related transcripts provides clues about new functionalities previously unreported. We hypothesized that the production of toxins in the Harderian gland is probably used for protection against pathogenic microorganisms, which might be helpful during prey neutralization.

5.2 Background

Within the most enigmatic organs of the vertebrates, the Harderian gland has been considered the last remaining large organ of widespread distribution, which its function has still to be cleared [311]. This gland is located within the eye's orbit (Figure 5.1) occupying a larger volume [309, 459-461] that in some species can even surpass in dimension the eye size itself [311]. The Harderian gland was firstly discovered in deer by Harder in 1964 [462], and since then it has been subsequently reported to occurs in most terrestrial vertebrates, such as reptiles, birds and mammals except bats, cows, horses, terrestrial carnivores and higher primates [311]. This gland seems to play an important role in the terrestrial environment, because it has not been found in fish, aquatic urodeles, nor in the aquatic forms of anurans; however, it is present in some secondary aquatic vertebrates such as crocodilians and cetaceans [311]. The proposed functions for the Harderian gland include: a source of saliva, pheromones, thermoregulatory lipids and growth factors, photoprotective organ, a site of immune response and osmoregulation [311, 312].

In Squamates, based on its anatomical position the Harderian gland has been associated with the nictitating membrane and ascribed as part of a retinal-pineal axis [311, 463-467]. Although, it was later well-established its morphological relationship with the vomeronasal organ (VNO) with a likely role in vomerolfaction and/or chemosensory [309, 310]. Several hypotheses have been discussed about the function of the Harderian gland in Squamates. In addition to the previously mentioned, the Harderian gland has been also considered an accessory salivary gland [468], which may function as a source of lubrication [459, 469] and as a production and secretion site of digestive enzymes [464]. However, apart from snakes and pygopods, the secretions of the Harderian gland empties its contents directly into the orbital space [464, 470], thus, seeming unlikely that the Harderian gland act as an accessory salivary in most of Squamates, as proposed [468]. Nonetheless, a deep revision revealed that there is little evidence suggesting a function in digestion [471]. On the contrary, the morphology of the Squamate Harderian gland and the presence of alternate secretory sources suggests it unlikely to play a role in the orbital lubrication [310]. There is evidence that the Harderian gland, which is the largest cephalic gland in garter snakes, plays a critical role transferring chemical signals (female pheromones, feeding cues) to the chemosensory epithelium of the VNO (unpublished work) [472]. Moreover, it was discovered that the capability to response against female pheromone decreased in male garter snakes after removal of their Harderian glands, thus affecting mating [472]. Thus, the most likely role of Harderian gland in Squamates could be related to chemosensory/vomerolfaction [310, 464, 471-473].



Figure 5.1. Two cephalic glands in *Caraiba andreae*, the Harderian gland (Hg) and the Duvernoy's gland (Dg). Picture was obtained during glands extraction.

This gland has been considered “as the most inconstant of all the glands of the snake’s head in the extent of its development”, since it varies from the condition of an incipient glandular tissue to an obvious organ in *Dryophis*, *Homalopsis*, *Enhydris*, or *Platurus* [474, 475], and also in the African egg-eating snake *Dasypeltis scabra* [476]. The Harderian gland is often very large in ryptozoic, fossorial [473, 477-479] and species that eat worms or slugs like *Tropidodipsas sartorii* [480]. The gland is present in sea-snakes but reduced in the most pelagic forms [460]. The assigned lubricant function of the Harderian gland in snakes is supported by the existing connection from the orbit around the eye to the mouth via the lachrymal duct [481, 482], which carries fluid from the Harderian gland (which lies medial to the eye) to the Jacobson’s organ [311]. It is noticeable that all the snakes examined by West [474] had a typical albuminous (serous) Harderian gland with much smaller alveoli than

those of the parotid, which originates the poison-gland in the venomous species. Indeed, there is limited information of what exactly the Harderian gland does in most snakes.

Next Generation Sequencing (NGS) approaches have become invaluable to characterize several tissues from multiple organisms, including snake's cephalic glands providing a complete picture of venom composition [69, 234, 247, 251-253, 255-257, 297, 299]. High-throughput sequencing approaches have been widely used to explore venom glands for drug discovery and therapeutic aims [69, 234, 298, 483]. In addition, *de novo* assembly of high-quality reads provided by methodologies based on Illumina sequencing technologies has been demonstrated to be promising, even in non-model animals like snakes [257, 484]. However, toxin-secreting oral glands in non-front-fanged snakes have been examined in a lesser degree [256, 297].

Herein, we characterized the Harderian gland's transcriptome from three colubrid snakes from Cuba with different ecological niches: *C. cantherigerus*, *C. andreae* and *T. variabilis*. The first two species are both terrestrial and diurnal active foragers, while the third one is a freshwater, nocturnal sit-and-wait predator [300]. The Harderian gland's transcriptome revealed transcripts related to binding/transporters, antimicrobial and immunological proteins, toxins as well as other components found in snake venoms. This evidence suggest that the role of this gland may be diverse, which could be linked to protection against possible harmful bacteria and maybe even to prey subdual. To our knowledge, the present study provides the first wide-transcriptome analysis of the Harderian gland in reptiles.

5.3 Methods

Specimens sampling and the Harderian gland extraction

Here, we sequenced the Harderian gland transcriptome of three species of colubrid snakes from Cuba: *Caraiba andreae* (Cuban Lesser Racer) (Ca), *Cubophis cantherigerus* (Cuban Racer) (Cc) and *Tretanorhinus variabilis* (Cuban Water Snake) (Tv). All specimens were sampled from the wild: Ca was sampled at Pico San Juan, Cienfuegos; Cc in Guanahacabibes, Pinar del Río; while Tv was sampled in Santa Fe, La Habana. Specimens weight was 44.9 g for Ca, 160.4 g for Cc and for 94.1 g for Tv, while size (a-snout-to-vent-length) was 382 mm for Ca, 537 mm for Cc and 323 mm for Tv. Unlike previous studies [252, 257], animals were not electrically stimulated before gland extraction. Each specimen was euthanized with sodium pentobarbital injection as described in [257] and both of their Harderian glands were removed (Figure 5.1). Then, glands were stored in a 2mL Eppendorf tube containing RNA stabilization reagent RNeasy (Thermo Fisher Scientific, Waltham, MA) and weighted in an analytical balance (model AW-224, Sartorius AG, Gottingen, Germany)

and the mass of each gland pair was determined by the difference with an Eppendorf containing 1mL of RNAlater without gland, respectively. All animals were manipulated according to the University of Havana Animal Care, based on the Institutional Animal Care and Use Committees (IACUCs) Guide [485]. Vouchers were deposited under the followed field code at the herpetological collections of the Museo de Historia Natural “Felipe Poey”, University of Havana (Cc and Tv, CHC-222 and CHC-224, respectively) and the Instituto de Ecología y Sistemática (Ca, CHC-225), La Habana, Cuba.

RNA extraction and Illumina sequencing

Total RNA extraction from each snake gland was carried out with Qiagen's RNeasy Mini kit (Venlo, The Netherlands). The extraction followed the protocol for purification of total RNA from animal tissues provided by the manufacturer. Briefly, both glands for each specimen (Ca: 87.4 mg; Cc: 82.1 mg and Tv: 91.3 mg) were transferred to a suitable vessel (Microwtube® Microcentrifuge Tube with Screw Cap, Simport Scientific) and added 600 µL of homogenized buffer (RLT buffer). Then, glands were disrupted and homogenated using Precellys® 24 tissue homogenizer (Bertin Technologies, Montigny le Bretonneux, France). The lysate was centrifuged at 8000× g for 3 min (Centrifuge – VWR Micro Star 17R, Radnor, Pennsylvania, US) and the supernatant was transferred to a new vessel where it was mixed with 70% ethanol (1:1). After this step, 700 µL of the sample was transferred to an RNeasy spin column, which was centrifuged for 15 s at 8000× g and the flow-through was discarded. Subsequently, 700 µL of RW1 buffer was added to the spin column, which was centrifuged as referred above. Afterwards, washing steps were continued by addition of 500 µL of RPE buffer with the centrifugation under the same conditions. Before elution, a last washing step was performed with 500 µL of RPE buffer and the column was then centrifuged for 2 min at 8000× g and the flow-through discarded. Finally, a total volume of 30 µL of sample was eluted using RNase-free water.

Before sequencing, the total RNA integrity was examined by agarose gel and then quantified using a master solution containing Quant-IT reagent (Invitrogen, Carlsbad, CA, USA) (1 µL × n samples) and Quant-IT working solution (199 µL × n samples) was prepared for quantification of the RNA content of the samples. Afterwards, 190 µL of master solution plus 10 µL of Quant-IT broad range RNA was mixed to prepare the standards used to quantify the sample RNA: tubes containing 1 µL of sample RNA and 199 µL of master solution were prepared, vortexed and incubated for 2 min at room temperature. Finally, RNA concentration was measured photometrically with Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA). RNA Integrity Number (RIN) as criteria of total RNA integrity was also evaluated with the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Then, total RNA extracted from each species (Ca: 11.47 µg, RIN: 8.2; Cc: 11.87 µg, RIN: 8.8 and Tv: 10.89

µg, RIN: 7.6) was used for library preparation on MacroGen, **Inc.** (Seoul, South Korea) using the TruSeq stranded total RNA library prep with Ribo-Zero Globin kit and finally sequenced in one lane on the Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA, USA) with 100-base-pair (bp) paired-end reads (PE). The Illumina-oligonucleotides adapter sequences used were as follow: Ca: TruSeq Adapter, Index 4 (5' GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGAC-CAATCTCGTATGCCGTCTTCTGCTTG); Cc: TruSeq adapter, Index 2 (5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAACGATGTATCTCGTATGCCGTCTTCTGCTTG); Tv: TruSeq Adapter, Index 7 (5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAACGATCATCTCGTATGCCGTCTTCTGCTTG) and the TruSeq Universal Adapter (5' AATGATACGGCGACCACCGAGATCT-ACACTCTTTCCCTACACGACGCTCTTCCGATCT).

Transcriptome assembly and bioinformatics analyses

Illumina data were analyzed using an adapted version of the workflow described at [247] which includes available NGS algorithms and in-house scripts. The first step to process NGS data was to trim and crop the paired-end Illumina raw reads as well as to remove adapters. This procedure was carried out by TrimmomaticPE (version 0.35) where some parameters were set to perform the trimming tasks and to keep the read quality. Afterwards, identical sequencing reads were collapsed using fastx-collapser (FastX Toolkit version 0.0.14) and resulting reads were considered as high-quality reads suitable for the following assembly process.

Due to the difficulty of a *de novo* transcriptome assembly, we performed a multi-assembler assembly approach, involving De Bruijn graph assemblers, as well as Overlapping Layout Consensus (OLC) assemblers combining results for a final data set prior to annotation. In this sense, the De Bruijn graph-based assemblers are more suitable for managing short read sequences as those obtained by an Illumina platform, while OLC assemblers used to be better dealing with larger sequences. Thus, collapsed sequences from the preprocessing step were assembled with Oases (version 0.2.08) [249], a *de novo* De Bruijn transcriptome assembler based on Velvet. Since the choice of the k-mer length (short sequence fragments representing nodes) is a subjective decision that will affect the recovery of longer transcripts or a larger transcript diversity, we performed the assembly process using an array of k-mer values, from 23 to 43 with a step value of 2. According to several metrics as N50, average contig length, total assembled nucleotides, maximum contig length, total number of contigs and number of singletons, several Oases assemblies were selected and merged giving rise to a subset of assembled sequences. It is worth mentioning

that as described previously the merging assemblies from different k-mer values performed best on the total number of assembled transcripts [486].

Collapsed sequences were also assembled with Trinity (version trinityrnaseq_r20131110) [248], an assembler that combines three independent software modules dividing sequence data into many De Bruijn graphs. A minimum length of 100 nucleotides long per contig was required. Finally, assembled sequences from Trinity and Oases were reassembled with CAP3 (Version Date: 08/06/2013) [250], an OLC assembler based on multiple sequence alignments. To this purpose a minimum 98% identity sequence was required to overlap nucleotide sequences.

The annotation process was carried out in two steps. First, given our previous knowledge of the interspersed repeats and low complexity sequences existence in transcriptome data, CAP3 contigs as well as CAP3 singletons were masked using RepeatMasker (version 4.0.5) [487]. The program screens DNA sequences for interspersed repeats and low complexity regions included in the Repbase database, a comprehensive database of repetitive element consensus sequences (version 19.08, update of September 2014) (<http://www.girinst.org>). Finally, masked sequences were blasted (blastall version 2.2.26) [488] against non-redundant nucleotide NCBI database (nt, update November 2014) specifying a cut-off value of $e-03$ and against Uniref100, a database that contains all Uniprot Knowledgebase records plus selected UniParc records.

Secondly, all CAP3 obtained contigs were imported into CLC Genomics Workbench (version 8.5.1) (Qiagen, Aarhus, Denmark) and used as references to map against the raw paired-ends reads. Then, sequences resulting as positive in the Blast search were concatenated with the CLC-mapped-contigs to estimate the expression using RPKM (Reads Per Kilobase per Million mapped reads) [489] and TPM (transcripts per million). In this work, the RPKM was the coefficient chosen to report and compare the estimated expression for each positive blast hit. Then, 25 keywords, selected from a survey over the 100 most expressed contigs were used to estimate the frequency of a protein/isoform occur in the list of the 1000 contigs with higher RPKM. The keywords used were as follow: ribosomal, uncharacterized, channel, venom, metalloprotease, mitochondrial, elongation, galectin-4-like, lipocalin-like, binding, toxin, actin, Bactericidal/permeability-increasing, phospholipase, cysteine-rich secretory protein, protease, oxidoreductase, ficolin, pentaxin, omanin, growth, cystatin-like, three-finger, nesprin-1 and interleukin-31.

GO annotation of non-toxins and non-characterized proteins

Non-characterized contigs were first analyzed using the Blast2Go software (version 3.3) blasting against the UniProtKB protein sequence database in the Metazoan section [204] and in the venoms section using BlastX algorithms, specifying a cut-off value of $e-03$

and BLOSUM62 as scoring matrix. Afterwards, non-toxins and non-characterized contigs were functionally annotated according to the Gene Ontology (GO) (<http://www.geneontology.org/>) nomenclature. Then, GO terms were assigned to each contig and annotated, per GO Distribution by Level (2), regarding the three major GO categories: Biological Process (BP), Molecular Function (MF), and Cell Component (CC).

Harderian gland toxin characterization

Snake Harderian toxin transcripts were selected from best blast-hit descriptions identifying GenBank entries belonging to the taxonomic suborder Serpentes. The subsets of sequences exhibiting similarity to snakes' sequences were further filtered using a list of keywords (including the acronyms of all known toxin protein families described so far) to distinguish putative snake venom toxins from non-toxin (ribosomal, mitochondrial, nuclear, etc.) ordinary proteins. In this sense, more than 160 keywords were used to classify blast-hit descriptions into several main toxin protein families. Putative miss-annotated sequences were also taken into consideration.

As described previously [247], the phylogenetically nearest amino acid full-length sequence was used as a reference onto which the six possible reading frames of toxin family-specific obtained nucleotide sequences translation exhibiting e-value thresholds better than e^{-03} were aligned to create a multiple alignment using Cobalt (version 2.0.2, build June 2013) [490]. The high number of mutations introduced during the sequencing process gave rise to high frameshift rate. Thus, multiple frames from the same contig could be contiguously aligned to the reference. For further analysis, we considered mutation-free full-length sequences as well as sequences with a single frame error-free aligned sequences to the reference.

Transcript abundance of a given toxin protein family was estimated by mapping pre-collapsed reads back to the coding sequences of previously annotated toxin-encoding transcripts. To this end, bowtie2 (version 2.2.4) [491] with `–very-sensitive` (i.e. `–end-to-end` mode) preset parameters and bowtie (version 1.0.0) with `–best` reporting parameters were used. The relative expression of transcripts resulted as Blast positive hits was calculated according to the RPKM (Reads per Kilobase of exon per Million mapped reads) standard procedure previously described [489].

In this study, the RPKM was chosen to estimate the relative expression of each toxin-transcript to facilitate the toxin expression level comparison with previous works. As previously cited in [252], the relative expression of each toxin protein family (mol %), was calculated as the number of reads assigned to this protein family (R_i) normalized by the length (in nucleotides) of the reference transcript sequence (ntREF) and expressed as the %

of total reads in the snake transcriptome (ΣReads): $\text{mol\% toxin family "i"} = \%[(\text{Ri}/\text{ntREF})/\Sigma\text{Reads}]$.

5.4 Results and Discussion

Harderian gland transcriptome and assembly statistics

We processed the Harderian gland's transcriptome of three species of colubrid snakes from Cuba using Illumina HiSeq2000 100 bp (base-pairs) paired-end (PE) with an average insert size of ~ 280 bp, generating a total of 54,593,038 of reads from *Caraiba andreae* (Ca), 50,762,546 from *Cubophis cantherigerus* (Cc) and 56,115,370 from *Tretanorhinus variabilis* (Tv). More than 5.5 Gb were identified in Ca and Tv while around 5.1 Gb resulted in Cc (Table 5.1). Finally, as displayed in Table 5.1, 63,333 contigs were obtained from *de novo* assembly in Ca, while 27,666 assembled contigs were obtained from Cc and 62,324 contigs from Tv. These statistics support that the workflow applied was a suitable strategy for *de novo* assembly using Illumina reads, although some metrics largely used in genomic analysis seem not to be appropriated to evaluate the quality of a transcriptome assembly [492]. Resulting Raw data sequences are archived under accession number SRP103723 in the NCBI Sequence Read Archive SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra> [492]. The accession codes by species are SRX2730337 for *Caraiba andreae*, SRX2730335 for *Cubophis cantherigerus* and SRX2730333 for *Tretanorhinus variabilis*.

Table 5.1. Statistics of the transcriptomes. The table shows a summary of the statistics for *de novo* assembly of the Harderian gland transcriptome of three Cuban colubrids: *Caraiba andreae* (Ca), *Cubophis cantherigerus* (Cc) and *Tretanorhinus variabilis* (Tv).

		Ca	Cc	Tv
RAW DATA	Read Count	54,593,038	50,762,546	56,115,370
	Total Bases	5,513,896,838	5,127,017,146	5,667,652,370
	GC(%)	51.78	48.8	52
	AT(%)	48.22	51.2	47.74
	Q20(%)	94.35	95.27	95.03
	Q30(%)	87.59	89.49	90.07
CAP3_assembly_stats	# contigs (>= 0 bp)	63,333	27,666	62,324
	# contigs (>= 1000 bp)	11,879	9,693	14,494
	# contigs (>= 5000 bp)	402	393	636
	# contigs (>= 10000 bp)	33	35	48
	# contigs (>= 25000 bp)	2	2	2
	# contigs (>= 50000 bp)	0	0	0
	Largest contig	29,430	29,155	28,545
	Total length (>= 0 bp)"all contigs"	46,223,254	29,666,985	52,264,819
	GC (%)	42.87	42.65	43.9
	N50	1092	1602	1337
	L50	10,568	5175	10,078
	# N's per 100 kbp	0.46	0.02	0.24
	# imported references (contigs)	63,333	27,666	62,324
CLC mapped reads	# number of reads (paired)	42,997,058 (78.76%)	41,732,724 (82.21%)	45,884,816 (81.77%)
	# reads mapped in pairs	32,702,780 (76.06%)	28,594,296 (68.52%)	30,750,932 (67.0)
	# reads mapped in broken pairs	4,529,254 (10.53%)	6,655,635 (15.95%)	5,139,839 (11.20)
	# reads not mapped	5,765,024 (13.41%)	6,482,793 (15.53%)	9,994,045 (21.78%)
Contigs matched as positive blast hits	total of positive hits	45,258 (71.46%)	23,370 (84.47%)	36,982 (59.33%)
	matched as potential toxins	1632	781	1299
	non-characterized within potential toxins	162	90	218
Total Bases: The total number of bases in reads identified				
Read Count: The total number of sequence reads				
GC(%): The GC percentage in sequence reads				
AT(%): The AT percentage in sequence reads				
Q20(%): The percentage of bases in which the phred score is above 20 (20 means a 1% chance)				
Q30(%): The percentage of bases in which the phred score is above 30				
# (number) N's per 100 kbp: could represent ambiguities in the graph behind the assembly				
N's: A "N" is designated in the sequence when the sequencer could not determine which base is present				
N50: average contigs length				
L50: number of contigs whose length is >= N50				
paired: there are two 'reads' in a pair				

Estimation of transcript expression

Assembled contigs were blasted (see Materials and Methods) resulting in 45,258 hits for Ca, 23,370 for Cc, and 36,982 for Tv, representing 71.46%, 84.47%, and 59.33% of the total contigs assembled, respectively. Out of total hits, 1,632, 781 and 1,299 contigs matched potential toxins transcript for Ca, Cc and Tv, respectively (Table 5.1). The assembled contigs were imported and used as reference for mapping against raw paired-ends reads, which yielded more than 40 million of reads in each species. Overall, represented 78.76% of mapped reads for Ca, 82.21% for Cc and 81.77% for Tv. In addition, 76.06% of those mapped reads were correctly paired in Ca while around 68.52% and 67.02% were paired in Cc and Tv, respectively. Moreover, around 4.5 million of reads were mapped in broken pairs representing 10.53% out of the total reads mapped in Ca, while were around 6.6 million in Cc (15.95%) and 5.1 million (11.20%) in Tv. Moreover, only 13.41% of the total reads sequenced were not mapped to reference in Ca, while in Cc and Tv represented 15.53% and 21.78%, respectively (see Table 5.1).

The resulting expression estimated for all contigs is shown in Supplementary Table 5.1. The values of the parameters Reads per Kilobase of Exon per Million Mapped Reads (RPKM) and Transcripts per Million (TPM) calculated showed a similar distribution, having both similar weights for discussion about contigs expression. Regardless the selection of the RPKM or TPM, it does not influence the interpretation of the transcript expression levels. We used RPKM to provide a relative expression of toxin-transcripts, to compare with previous works, considering similarities in both the applied assembling methodology [252] and species taxonomy [256].

In terms of RPKM expression, it is common the use of most abundant contigs for accurate analyses, as a way of figuring out their relative expression. In this study, we used the first 25 contigs, since the 25th contig value represented approximately 1% of the maximum value of the RPKM calculated (Figure 5.2A) as a general trend for the snakes under study. This distribution suggests that some elements like these 25 MEC, may have a high expression relatively to the other transcripts of the gland. That is why these contigs were firstly analyzed and several keywords extracted from its respective description were used as a guide for further analyses. Products related to these contigs should have an important role for the gland function or simply represent constitutive proteins. However, other transcripts with less expression can also be relevant within the final products of gland secretion.

The expression of the other contigs from Contig 25 to Contig 1000, decreased gradually (Figure 5.2A). The value of RPKM of the Contig 1000 represented 0.01% of the maximum value of the RPKM, but represented just the 2.5% of the value of the Contig 25. Moreover, the Contig 5000 showed a RPKM value that represented 25% of the Contig 1000.

These elements supported the workflow used in the analysis of the contig expression since the expression value remained almost constant from Contig 1000 to Contig 5000 (Supplementary table 5.1). Unlike the sudden decline of the RPKM until Contig 25, the value of RPKM of Contig 5000 decreased by a quarter of the value of Contig 1000. Previous analyses provided reliability in profiling the transcriptome expression by following a standardized protocol. According to the RPKM, the expression profile was characterized by a high-level expression of the first 25 contigs, while most of the contigs from Contig 5000, showed similar expression. However, they should not be underestimated, since the expression level does not differ so much from the previous contigs.

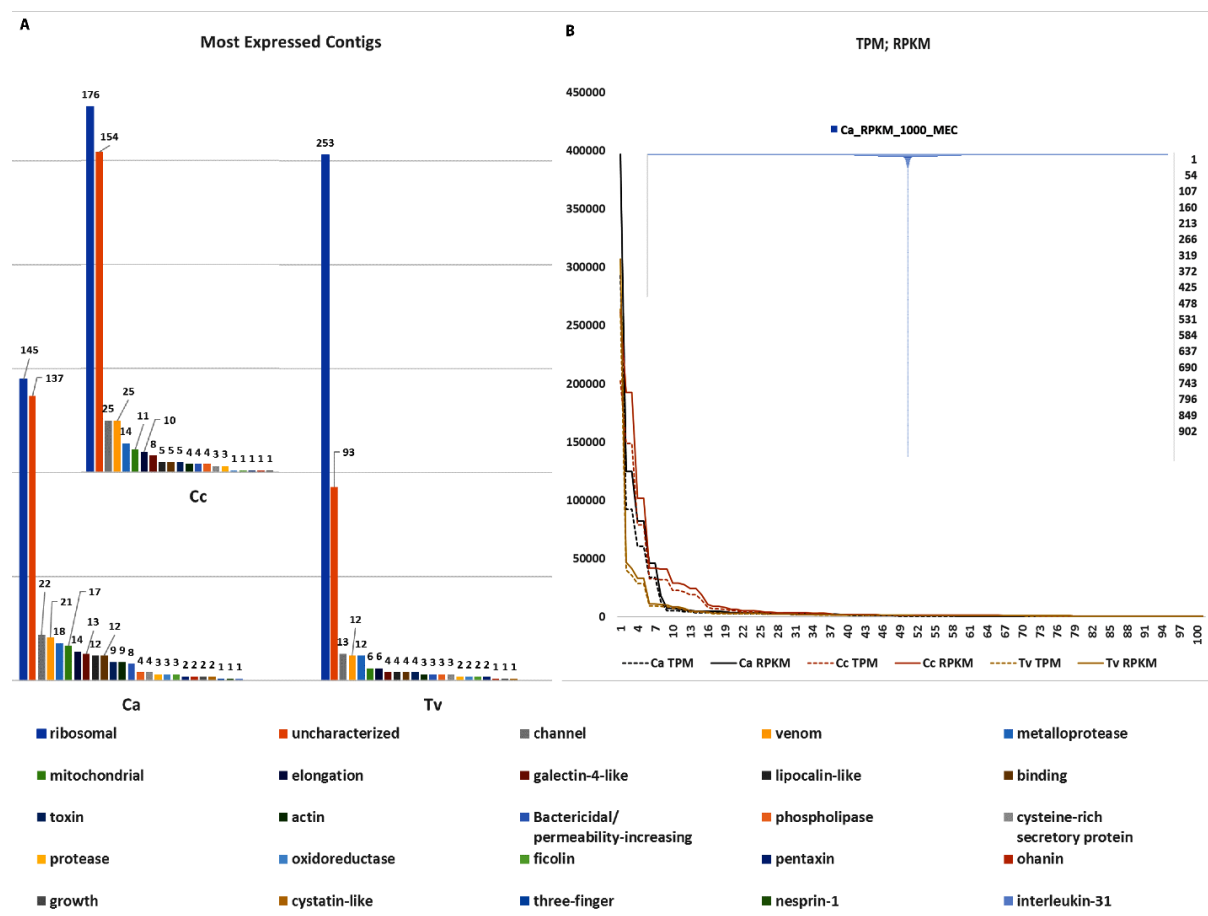


Figure 5.2. Most expressed contigs. Figure in panel A shows the 25 keywords used to map the relative frequency of the most expressed transcripts. In panel B is shown the distribution of RPKM among the first 1000 contigs sorted by expression level.

Most expressed transcripts

After inspecting the description of the 1000 MEC, 25 keywords were selected to estimate the occurrence of some proteins in the list of the 1000 MEC. Then, these 25 keywords were used as an additional filtering round for putative isoforms of each related protein. They matched a total of 465 hits for Ca, 459 for Cc and 434 for Tv within the 1000 MEC, representing around the 45% of the total of 1000 MEC surveyed. The two most

matched keywords in the three species were “uncharacterized” and “ribosomal”, which represented 31%, 33% and 21% of the total hits on Ca, Cc and Tv, respectively. It is noteworthy the high occurrence of uncharacterized proteins, representing an evidence of the limited knowledge about the components expressed in the Harderian glands. On the other hand, ion-channels-related transcript, binding-related and mitochondrial proteins were detected within the most expressed genes. However, some toxins-related transcripts were detected using the keyword “toxins” and “venom”. In the case of Ca, the keyword “venom” matched 21 times (in some cases two hits represented the same contig), while the keyword “toxins” matched 9 times. On the other hand, the keyword “toxins” matched 10 times in Cc and 4 times in Tv, while the “venom” term does not have any match in Cc and only 3 hits in Tv.

Other non-toxin contigs detected with high expression but with less frequency were pentaxin, nesprin-1, interleukin-31 and growth factors. Some contigs matched for the same protein within the 25 MEC list, suggesting the presence of isoforms that could play significant role in the Harderian gland function. Among them, could be mentioned lipocalin/lipocalin-like-toxins, bactericidal/permeability-increasing protein-like, cystatin-like, and some uncharacterized proteins, which were common for the species studied (Table 5.2).

Table 5.2. Description of the 25 most expressed contigs. The table shows the hits description of the 25 most expressed contigs.

Ca		Cc		Tv	
Contig name	Brief description	Contig name	Brief description	Contig name	Brief description
Contig56439	<u>non-characterized (a)</u>	Contig16771	common carp genome, scaffold: LG40, chromosome: 40	Contig28001	<u>non-characterized (a)</u>
Contig31378	<u>lipocalin-like toxin (b)</u>	Contig6200	<u>non-characterized (a)</u>	Contig58025	non-characterized
Contig31344	28S ribosomal RNA	Contig10455	<u>lipocalin-like toxin (b)</u>	Contig7413	putative helicase
Contig31377	<u>Lipocalin (c)</u>	Contig2534	internal transcribed spacer 2	Contig30972	<u>lipocalin-like toxin (b)</u>
Contig14	non-characterized (d)	Contig3414	non-characterized	Contig37414	<u>oxoglutarate dehydrogenase/ (d)</u>
Contig31260	<u>bactericidal/permeability-increasing protein-like (e)</u>	Contig5650	<u>Lipocalin (c)</u>	Contig7414	Plasminogen
Contig5204	<u>Ac1147-like protein (h)</u>	Contig17830	non-characterized	Contig22800	Caprin-1
Contig46534	non-characterized	Contig1453	non-characterized	Contig36433	mitochondrion (mtDNA)
Contig40910	<u>cystatin-like (f)</u>	Contig6201	non-characterized (d)	Contig23885	<u>Lipocalin (c)</u>
Contig52250	mitochondrion (mtDNA)	Contig3405	collagen alpha-1(I) chain-like	Contig33529	non-characterized
Contig36887	cohesin/phosphatase and actin regulator 4	Contig36	<u>lipocalin-like (c)</u>	Contig39200	ribosomal protein
Contig16918	non-characterized	Contig16886	common carp genome scaffold 000004095	Contig9843	Nuclear prelamin A recognition factor

Contig28866	NADH-ubiquinone oxidoreductase	Contig15750	<u>cystatin-like (f)</u>	Contig42716	extensin-like
Contig16942	vegetative cell wall protein gp1-like	Contig25213	non-characterized	Contig32435	<u>bactericidal/permeability-increasing protein-like (e)</u>
Contig31368	<u>non-characterized (g)</u>	Contig6774	<u>bactericidal/permeability-increasing protein-like (e)</u>	Contig53662	Rnase P RNA gene
Contig16891	serum amyloid P-component-like/ Pentaxin	Contig11807	<u>non-characterized (i)</u>	Contig81	Pentaxin
Contig42868	Ohanin	Contig3454	non-characterized	Contig67	non-characterized
Contig16916	<u>non-characterized (g)</u>	Contig15	non-characterized	Contig43635	inhibitor of growth family
Contig3115	Reticulon	Contig9428	internal transcribed spacer 2	Contig5	non-characterized
Contig16880	elongation factor 1-alpha	Contig21470	<u>non-characterized (g)</u>	Contig25982	Glycine proline rich protein
Contig46973	internal transcribed spacer 2/ zinc finger protein 345-like	Contig25071	<u>bactericidal/permeability-increasing protein-like (e)</u>	Contig61351	parvalbumin beta-like
Contig46974	<u>Ac1147-like protein (h)</u>	Contig9363	mitochondrion (mtDNA)	Contig25983	Beta-keratin
Contig33421	<u>lipocalin-like (c)</u>	Contig3411	clone 48B complement C3-2 precursor/RNA binding	Contig21602	non-characterized
Contig16922	<u>non-characterized (i)</u>	Contig3422	<u>cystatin-like (f)</u>	Contig26051	non-characterized
GenBank /UniProt Accession Number Uniref: a (UPI00051ABB17); b (JQ354966.1/H9CNK5_ERYML); c (JQ340879.1/I0BWR4_DISTY); d (-/7F1M8_PELSI); e (AM774031.1/C6S3P7_DEIAC); f (-/UPI0004418F0E); g (XM_007442884.2/UPI0004423762); h (-/J3SBR0_CROAD); i (XR_001560236.1/-)					

Putative snake's toxins were found into the most expressed contigs. In detail, 14 different contigs were identified as venom metalloproteinases (SVMPs) in the description of the 1000 MEC of Ca, while none contig in Cc and only 2 contigs in Tv were detected. Besides, lipocalin-like toxins were found 5, 3 and 5 times in Ca, Cc and Tv, respectively.

In this sense, lipocalins constitute a heterogeneous group of secreted proteins that bind a wide variety of small hydrophobic ligands with high affinity [493]. This protein family has been ascribed to be involved in many biological processes including immune response, olfaction and pheromone transport [493-495]. However, some of these proteins are unlikely to be involved in olfaction, because they have shown a weak sequence similarity with a high expressed lipocalin in a specialized tissue from a frog [496]. Moreover, lipocalin has been proposed as venom or at least as important component of the venom [497], since it has been found in gland secretions of hematophagous invertebrates and vertebrates as well [2]. In snakes, lipocalin was discovered as a novelty in a transcriptomic study of venom glands from a variety of advanced snakes [253]. Indeed, lipocalin was reported in *Atractaspis aterrima* venom gland transcriptome as one of the most expressed transcripts [498]; still more in the Duvernoy's gland transcriptome from the colubrid *Oxyrhopus guibei*, representing around 29% of the sequencing reads [256].

Further work might be done to confirm whether lipocalin could play a role just in binding/mediated-olfactory, in immune-response or as a toxic component. Cysteine-rich secretory proteins (CRISP) were identified in two different contigs for Ca, one for Cc and two for Tv, while three-finger toxin (3FTx) was detected three times in Cc, twice in Ca and none in Tv. Moreover, the promoter region of phospholipase A2 (PLA2) was also detected within the 1000 MEC. On the other hand, phospholipase A2 inhibitor was found in the list of the most expressed and with relative high frequency (Supplementary Table 5.1). Cystatin/cystatin-like and protein disulfide-isomerase (PDI) were found among the most expressed transcripts, which are frequently found in venoms [253, 257, 499]. Pentaxin (pentraxin) was also detected as one of the most representative genes in the three species, which function has also been related to acute immunological responses [500]. Other elements showing high expression level were elongation factors, which were detected within the 25 MEC and matched with some of the 1000 MEC with high RPKM values. These elongation factors, called as the “workhorses of protein synthesis on the ribosome,” play an important role in the elongating of the new polypeptide chain [501].

Finally, some transcript hits to “galectin-4-like” were detected with relative high frequency in the 1000 MEC. This protein, comprises a tandem-repeat galectin with two distinct carbohydrate recognition domains [502], and has been strongly associated with binding/delivery processes [503-505], mainly for the alimentary tract functioning as a stabilizer of lipid raft and adherents junction [502]. It is noteworthy that galectins are basically lectins, also classified as carbohydrate binding proteins (CBP) being an important cell death-inducing ligands [506]. In this sense, some lectins like C-type lectins (CTL), have been so far identified as toxic components from several animals [2, 507, 508]. Though galectins and CTL seem to be unrelated, galectins have been proposed as a potential venom component since they were found in the blue-spotted stingray *Neotrygon kuhlii* at transcriptomic and proteomic level [499]. In the latter study, the galectin resulted as one of the most expressed genes and the second ranked in the proteomic analysis. Moreover, the proposed function of galectin as potential toxin or at least as a venom component is well supported by the previous evidence plus its capability of triggering the apoptotic cascade and inducing cell death [506]. Another finding of galectin occurred in the Duvernoy’s gland transcriptome of the colubrid *Oxyrhopus guibei*, where a transcript matched as galectin-4 occurred as one of the most expressed genes [256].

In general, the MEC found in the transcriptome of the Harderian gland suggests a likely role in the production of toxins and compounds involved in transport as ligand and even defense. The presence of bactericidal proteins in addition to snake’s toxins, shed light about possible involvement in defense against pathogens. The family of bactericidal/permeability-increasing proteins has been associated with the protection against

gram negative bacteria in human and mice [509], showing an evolutionary conservancy (due to their very important role) among very distinct mammals that could be occurring in other vertebrates, like reptiles. Both lipocalin and bactericidal/permeability-increasing protein were reported as overexpressed in the human and mice olfactory epithelium transcriptome [510], which make sense since this epithelium is continuously exposed to pathogen invasion [509]. Besides, the detection of some transcripts related to protein transport, suggests a likely role in vomerolfaction, seemingly related to the transport of pheromones or odorants as previously reported [472]. [472]. Besides, we detected transcripts related to olfactory-receptor among the 1000 MEC, but not within the higher expression level transcripts. In addition, other transcripts related to the synthesis of pheromones like estrogen receptor were not detected among the MEC, even within in the mentioned 1000 MEC. The presence of estrogen could be used as important signal of pheromones expression, since estrogen activates the production of female pheromone in adult red-sided garter snakes [511]. Other transcripts expression related to “binding/transporting” proteins, involved in the olfaction pathway like “vomeronasal receptors” [512], “odorant-proteins” [509, 510, 513], “retinol/retinol-binding proteins” [496], “neuropeptides” [510], were absent or detected with less RPKM value. However, at the transcriptomics level these evidences are weak to draw conclusions, since the presence of some compounds and its biological function are versatile, thus resulting ambiguous.

Non-toxin contigs Gene Ontology

Non-toxin and uncharacterized contigs were queried using Blast2Go software against non-redundant NCBI database [228]. Overall, we analyzed 24,469 non-toxin contigs for Ca, 14,386 for Cc and 24,114 for Tv (Figure 5.3A). Of these, 8,957 sequences for Ca, 5,055 for Cc and 8,569 for Tv resulted as blast hits using “snakes” as a taxonomy clade filter. Out of the total of contigs blasted, 2,025 sequences for Ca, 972 for Cc and 2,222 for Tv resulted as positive hits. In addition, we analyzed a total of 2,846 with mapping in Ca, 1,517 for Cc and 2,701 for Tv, while with GO annotation we obtained 10,641, 6,842 and 10,622 sequences for Ca, Cc and Tv, respectively (Figure 5.3A). The top-five species matched in the B2GO analysis (The BLAST top-hits species distribution) were *Tamnophis sirtalis* (8,277 sequences for Ca, 4,916 for Cc and 8,038 for Tv), *Protophthrops mucrosquamatus* (6,084 sequences for Ca, 3,784 for Cc and 6,066 for Tv), *Ophiophagus hannah* (4,269 sequences for Ca, 2,316 for Cc and 4,127 for Tv), *Python bivittatus* (3,238 sequences in Ca, 2,074 for Cc and 3,223 for Tv) and *Protophthrops flavoviridis* (374, 195 and 232 sequences for Ca, Cc, and Tv, respectively) (Figure 5.3B).

Contig sequences uncharacterized by annotation comprehended 162, 90 and 218 for Ca, Cc and Tv, respectively (Figure 5.5A). However, it was possible to annotate most of

those non-characterized proteins analyzed against the UniProtKB protein sequence database in the Metazoan section. Overall, 18, 8 and 38 contigs for Ca, Cc and Tv, respectively, remained without hits (Figure 5.5). Afterwards, it was possible to obtain the GO Annotation term for 62, 33 and 78 of the total of non-characterized proteins from Ca, Cc and Tv, respectively (Figure 5.5A). As displayed in Figure 5.5B, most of the non-characterized proteins matched were from the species *P. flavoviridis*, *O. hannah* and *T. sirtalis*). Finally, the GO terms were assigned to each annotated contig resulting in high similarities to the three species (Figure 5.6).

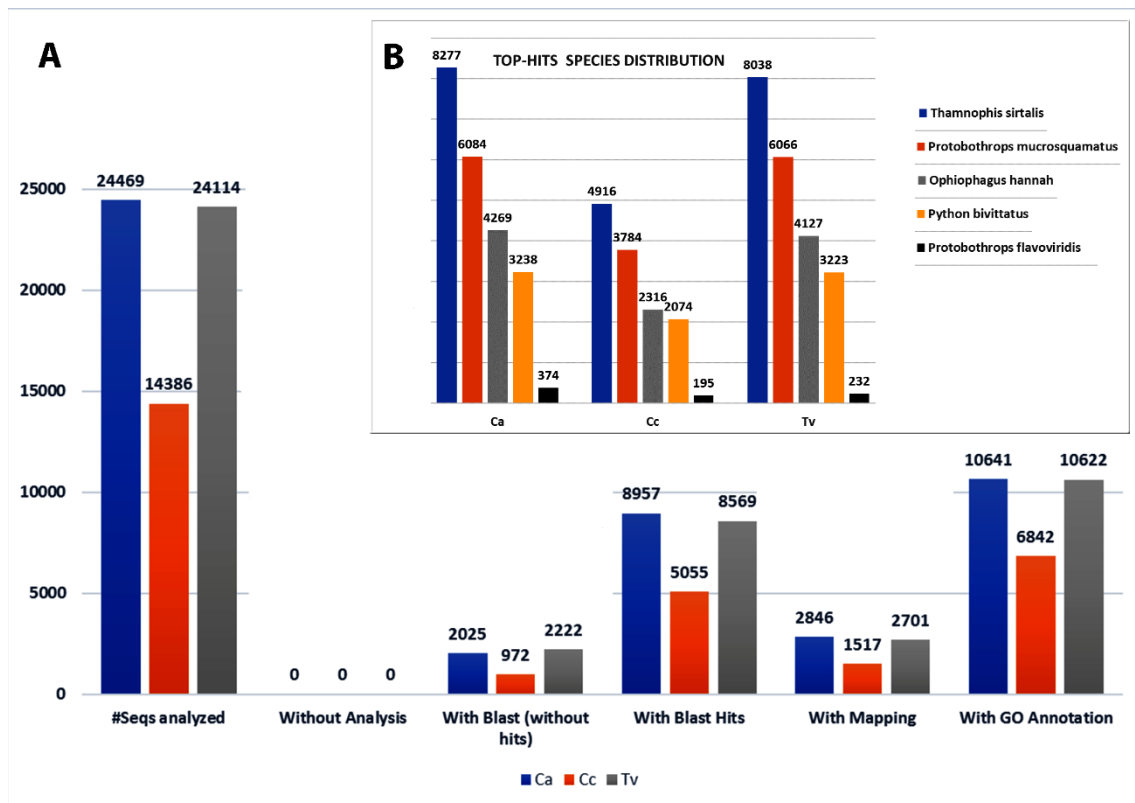


Figure 5.3. Gene ontology (GO) statistics of non-toxin contigs.A. Summary of blast and mapping statistics, B. top-hits distribution per number of hits retrieved for each species.

Then, GO terms were assigned to each contig and annotated according to GO Distribution by Level 2, regarding the three major GO categories: Biological Process (BP), Molecular Function (MF), and Cell Component (CC). Specifically, a total of 22,140, 11,301 and 10,164 of non-toxin contigs were associated to the categories BP, MF and CC in Ca, respectively. The total of hits for the species Cc and Tv in the category of BP were 14,317 and 22,234 in this order. In the categories of MF and CC were 7,329 and 6,680 for Cc, respectively; while for Tv were 11,184 and 10,048 of hits, in the same order of categories. As displayed in Figure 5.4 the distribution of hits for the category of BP, MF and CC was similar

in the three species in both groups analyzed “non-toxin” contigs (Figure 5.3, 5.4) and “non-characterized” proteins (Figure 5.5, 5.6).

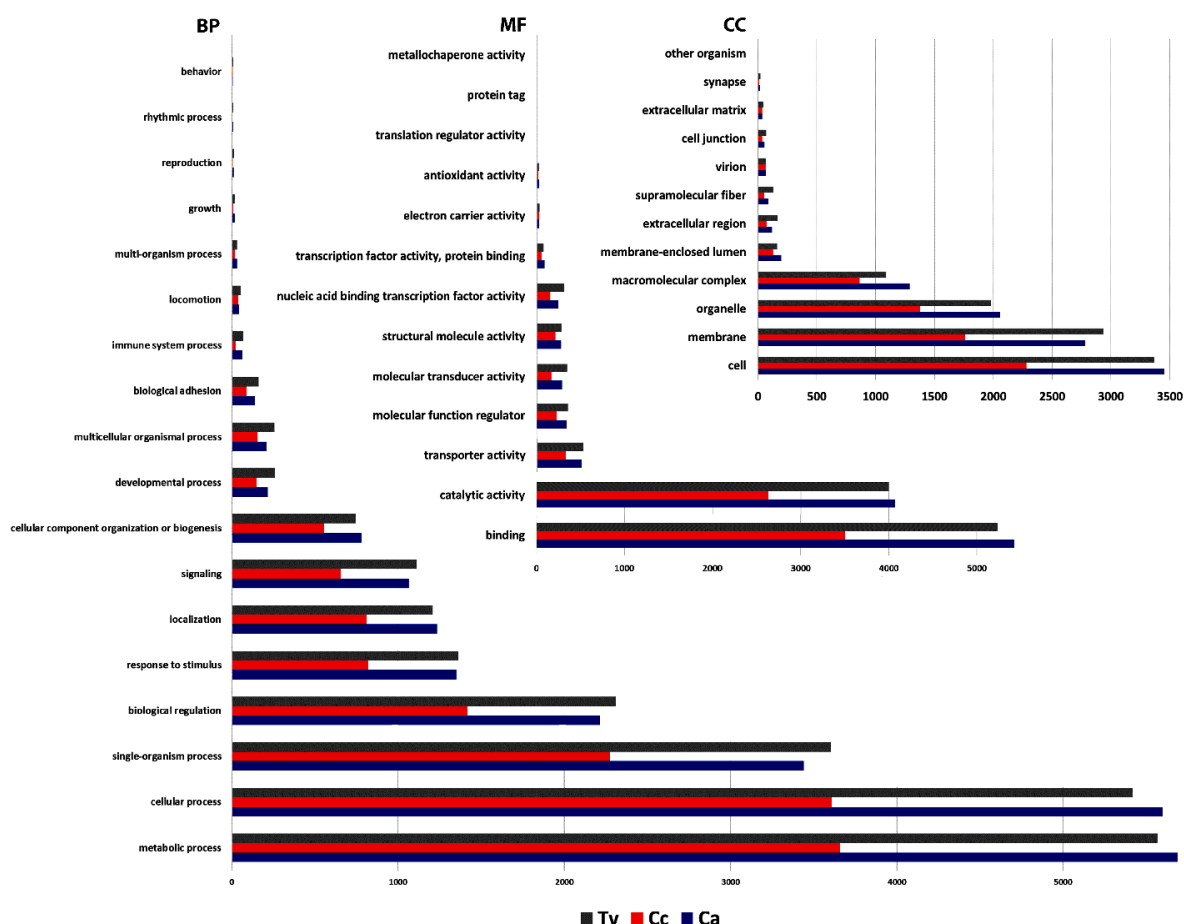


Figure 5.4. Gene ontology (GO) annotation of non-toxin contigs by level 2. Non-toxin contigs annotated regarding the three major GO categories: Biological Process (BP), Molecular Function (MF), and Cell Component (CC).

The first three GO terms in the category of BP were metabolic process (GO:0008152), cellular process (GO:0009987) and single-organism process (GO:0044699). In the others two categories, the first three GO terms were binding (GO:0005488), catalytic activity (GO:0003824) and transporter activity (GO:0005215) for MF, while for CC were cell (GO:0005623), membrane (GO:0016020) and organelle (GO:0043226). In general, the category of MF is useful to provide an overview about the function of the proteins or entities present in a tissue. It is noteworthy that the GO terms “binding” and “catalytic activity” were notably greater than the third GO term “transporter activity”. The GO term “binding” are related to non-covalent interaction of a molecule with one or more specific sites on another molecule, while “catalytic activity” involves catalysis regarding enzymes activity [225, 514]. However, the GO term “olfactory receptor activity” (GO:0004984, Synonyms: odorant

receptor activity) [514, 515] was absent, rather than expected for a tissue where the “transport of pheromone” or “odorants”, constitutes the main function, as proposed [472]. On the contrary, similar results were obtained for non-toxin transcripts with the BLAST2GO software in the venom-gland transcriptome of the diamondback rattlesnake *Crotalus adamanteus* [257], where “binding” and “catalytic activity” were the most significant GO terms obtained.

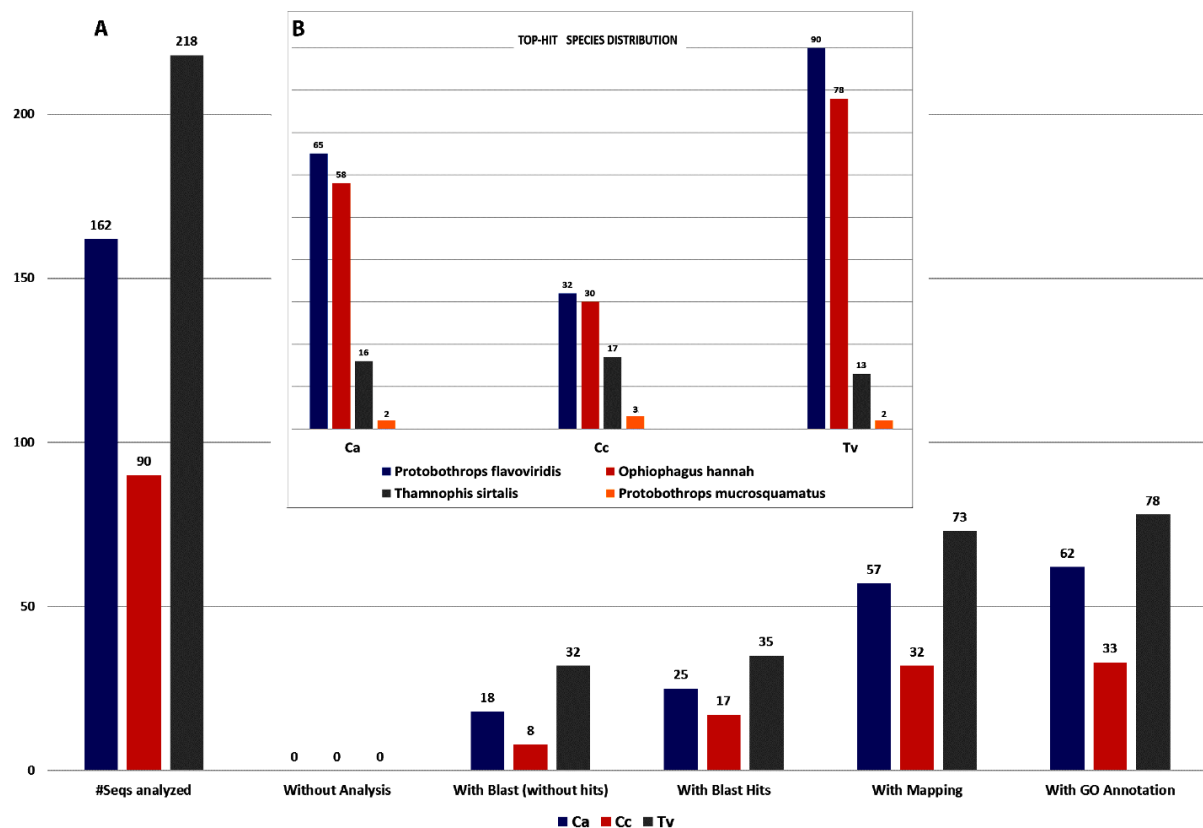


Figure 5.5. Gene ontology (GO) annotation of non-toxin contigs by level 2. Non-toxin contigs annotated regarding the three major GO categories: Biological Process (BP), Molecular Function (MF), and Cell Component (CC).

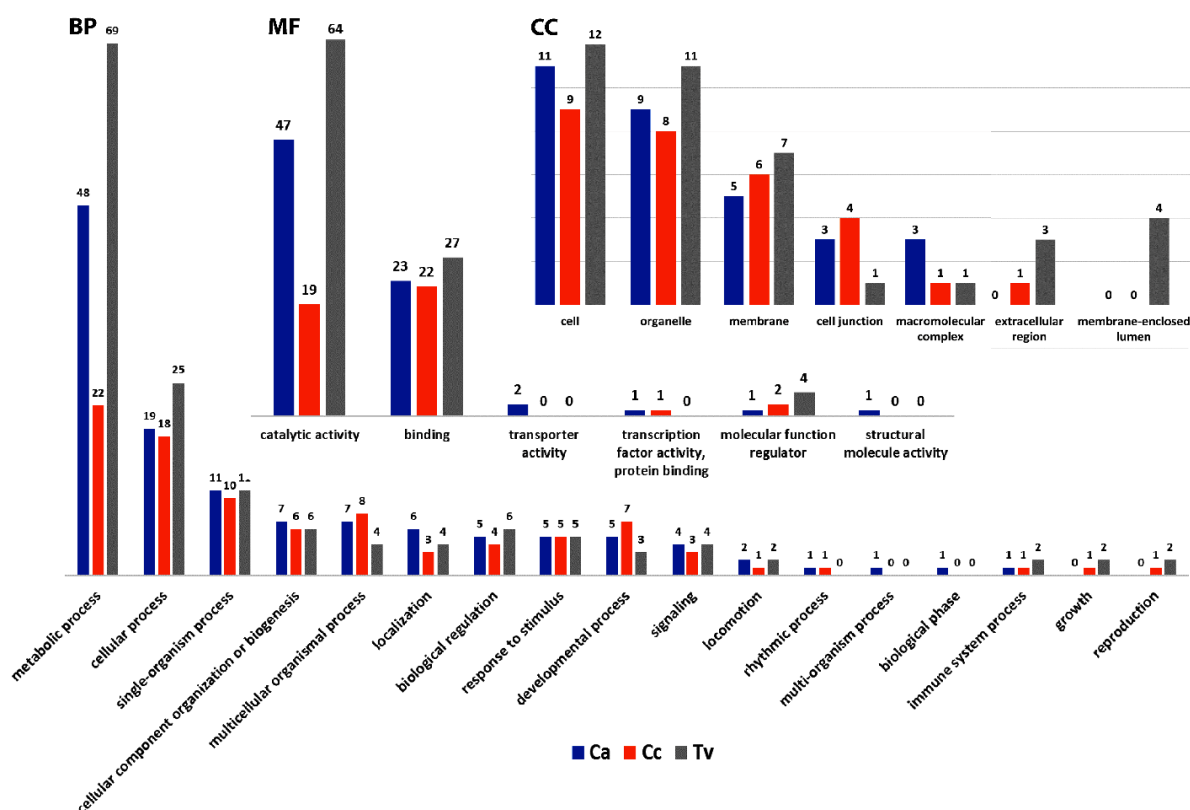


Figure 5.6. Gene ontology (GO) annotation of non-characterized proteins by level 2.

Harderian gland' toxin composition vs snake venom gland compositions

Given the fact that some toxins largely described belonging to the snake venom gland were identified by blast, we carried out a deeper analysis of the Harderian's gland toxin composition. We followed an adapted version of the data analysis protocol previously described [256].

First, we separated putative toxins from non-toxin transcripts. Overall, 1,632 contigs for Ca, 781 for Cc and 1,299 for Tv, were carefully analyzed as putative toxins (Table 5.1). The relative expression of snake's toxins of the three species is shown in Figure 5.7. The most expressed classes of snake's toxins in Ca were metalloproteinase SVMP (43.2%), CRISP (38.8%), Waprin (6.08%), Kunitz-type protease inhibitor (6.08%) and CTL (2.6%). Other toxin classes detected in less proportion were: 5'-nucleotidase (5-NT), Bradykinin potentiating peptide (BPP), Vascular endothelial growth factor (VEGF), Serine proteinase (SP), Ficolin (FIC), Hyaluronidase (HYA) and Phospholipase A2 (PLA2) (Table 5.3).

On the other hand, the most expressed snake's toxin in Cc was KUN amounting 44.8%, followed by WAP that represented 27.1% of the total of the detected snake's toxins. Surprisingly, FIC accounted for 9.9%, while 5-NT and CTL resulted in 5.0% and 4.3%, respectively. Besides, SP, SVMP and PLA2 were also detected, representing 2.8%, 2.2%

and 1.9% in this order. The classes with less expression found in Cc were VEGF (1.5%) and HYA (0.4%). Unlike Ca and Cc, the most expressed toxins in Tv was Three-Finger Toxin (3FTx) which resulted in 36.6% of relative abundance. In addition, CRISP resulted in 20.5%, although FIC resulted similar representing 18.3% of the total. Other classes detected in Tv with significant expression level were CTL (11.9%) and SVMP (8.3%). Besides, others toxin families also found with less relative expression level were: KUN, PLA2, VEGF, 5-NT, SP, WAP, BPP and HYA (Table 5.3). Despite some transcripts matched as ohanin/vespryn among the 1000 MEC, no reads could be mapped against it or did not align properly to the reference, which is in line with previous described results [256]. In addition, others transcript related to “snake venom proteins” such as: Nerve/Neurotrophic growth factors (NGF), Glutaminyl-peptide Cyclotransferase, L-amino acid oxidase (LAO), Cobra Venom Factor (CVF), Crotonamine, Crotoxin, Calglandulin and Sarafotoxin were also identified by blast means, but reads did not match with any translated ORFs. The fact of finding potential toxins among blast-hits could be associated to non-translated regions (mostly microsatellite DNA) or pseudogenes, as previously reported [252].

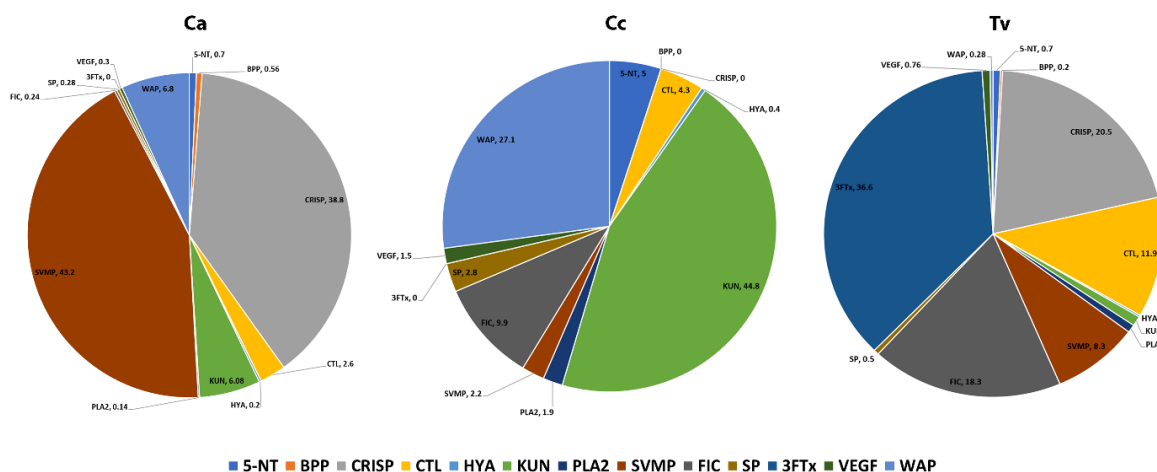


Figure 5.7. Relative expression of snake's toxins detected in the transcriptome of the Cuban colubrids: *Caraiiba andreae* (Ca), *Cubophis cantherigerus* (Cc) and *Tretanorhinus variabilis* (Tv).

In the above-mentioned study eight snake venom gland transcriptomes from five different taxa of Costa Rican (8-CRS) snakes were investigated using 454 pyrosequencing. The comparative analysis revealed taxon specific trends regarding venom composition. Given the wideness of snake venom gland toxin expression values obtained from those 8-CRS, we considered a RPKM average (AVE) value for each toxin family (Table 5.3), previously reported in [252] giving rise to weight specific toxin expression comparison between venom gland and Harderian gland. Firstly, the two most expressed toxins in Ca were compared. In the case of SVMP, the RPKM represented around 20% of the RPKM-

AVE, but seven times more than the resulting RPKM minimum value (RPKM-MIN) (different of zero) from the eight Costa Rican species. On the contrary, CRISP, the second family showing the highest relative expression resulted 1.8 times greater than RPKM-AVE of the 8-CRS, and represented 88.8 % of the maximum value of RPKM (RPKM-MAX) (Supplementary Table 5.2) suggesting that CRISP could play a more remarkable role in colubrid Harderian gland content than SVMP.

Table 5.3. Relative expression of snake Harderian gland toxins. Relative expression of the toxin classes found in the transcriptome of the Harderian gland of Cuban colubrids, and the average of the expression values calculated from previous published data [252].

snake's toxins	Ca		Cc		Tv		8 Costa Rican Species (Durban <i>et al.</i> ; 2011)	
	RPKM	%	RPKM	%	RPKM	%	RPKM-AVE	RPKM-MIN
5-NT	5422.9	0.7	103,439.0	5	9421.9	0.7	23,530.8	1510
BPP	4142.3	0.56	0	0	2754.3	0.2	1,339,871.9	188,484.5
CRISP	285,476.2	38.8	0	0	275,376.0	20.5	154,581.9	3693.4
CTL	19,121.4	2.6	87,780.2	4.3	160,848.3	11.9	249,381.9	13,071.8
HYA	1531.4	0.2	7621.3	0.4	2424.9	0.18	2750.1	658.9
KUN	44,708.7	6.08	921,848.5	44.8	13,170.4	0.98	10,322.1	4581.9
PLA2	992.8	0.14	39,402.4	1.9	10,613.2	0.79	1,873,990.1	344,017.0
SVMP	317,547.2	43.2	45,861.9	2.2	110,863.7	8.3	1,622,115.9	41,772.9
FIC	1763.4	0.24	203,263.9	9.9	245,854.7	18.3	N/D	N/D
SP	2053.1	0.28	58,348.1	2.8	6755.7	0.5	1,455,785.7	132,116.4
3FTx	0	0	0	0	491,571.0	36.6	361,111.1	147,008.54
VEGF	2221.9	0.3	31,076.8	1.5	10,252.1	0.76	N/D	N/D
WAP	50,003.2	6.8	558,997.3	27.1	3838.4	0.28	11,888.4	95,107.4
cuban colubrid species: Ca: <i>Caraiba andreae</i> ; Cc: <i>Cubophis cantherigerus</i> and Tv: <i>Tretanorhinus variabilis</i>								
snake's toxins; name of each toxin family detected: 5-NT: 5'-nucleotidase; BPP: Bradykinin potentiating peptide;								
CRISP: Cysteine-Rich Secretory Peptide; CTL: C-type lectin-like protein; HYA: Hyaluronidase; KUN: Kunitz-type inhibitor;								
PLA2: Phospholipase A2; SVMP: Snake venom metalloproteinase; FIC: Ficolin; SP: Serine Proteinase; 3FTx: Three-Finger Toxin;								
VEGF: Vascular endothelial growth factor; WAP: Waprin								
N/D: toxins family non-detected								
%: represents the relative expression between the toxins family detected								
RPKM-AVE: Average of RPKM of all toxins detected in Durban <i>et al.</i> ; 2011								
RPKM-MIN: Minimum of the RPKM detected (different of zero) among the toxins detected in Durban <i>et al.</i> ; 2011								

On the other hand, KUN showed a higher expression in the Ca, Cc and Tv species than the RPKM-AVE of the 8-CRS. In fact, KUN was 20 times more expressed in Cc than the RPKM-MAX reported for the Costa Rican species. In colubrid species, these toxins have been considered as one of the most relevant within the category of minor snake venom components [256]. Surprisingly, RPKM AVE of WAP on 8-CRS represented the 50% of the

RPKM value in Cc. In fact, WAP percentage toxin expression from Cuban snakes ranged from 0.28% to 27.1% while it was only identified in the Caribbean population of *Bothrops asper* (0.09% of venom content) being absent in the Pacific population of the same species. Interestingly, toxic effects of Cc bites have been documented in human. Inflammation, flushing, fever [516, 517] and even tissues necrosis were reported (unpublished). Besides, a death produced presumably by envenomation was reported after Cc bit a lizard (*Anolis equestris*) and in the same study was reported that Cc is opisthoglyphous [518], common condition for snakes non-lethal to humans [254].

3FTx transcripts were the most expressed transcripts within the snake toxins detected in the species Tv. When compare its RPKM with the RPKM-AVE of 8-CRS, resulted around 1.5 greater. Therefore, these proteins may play a prominent role in the Harderian gland. As largely described, they constitute the main venom components of snakes from the family Elapidae [519], but are also present in different subfamilies of Colubridae [256] where they have been frequently detected as highly abundant components with taxon-specific activities [520, 521]. As mentioned above, CRISPs were found as the second ranked at the relative expression level in Tv, representing 1.8 times greater than the RPKM-AVE of 8-CRS. CRISPs seem to be ubiquitous venom components, since have been found in almost all snake species, including colubrids [256]. Its function remains unclear, but must be relevant as toxic component in some snakes, because positive selection on CRISPs from Colubridae and Viperidae have been detected, unlike other reptiles and CRISPs mammals [522]. For some transcripts like lipocalin/lipocalin-like, cystatin and other common cellular components, the relative level of expression among the toxins family detected was different in each species. This finding is probably related to the ecological niche of each snake studied.

Overall, the production of these toxins in the Harderian's gland should not be underestimated, since this phenomenon has been reported repeatedly in the oral secretion of some nonvenomous species of snakes and lizards [2, 105, 306, 307, 523, 524]. In Squamates, some tissues can produce toxin-transcripts, because of the proposed single early origin of toxin encoding-genes, at the base of the Toxicofera [525, 526]. Moreover, this feature has not been only assigned to those organs or glands, which its embryological development is related to oral secretions. In addition to salivary and scent glands expression profiles from skin genes from gekkonid lizards and some snakes have revealed co-expression of toxin-related genes at comparable level found in snake venom gland [525]. Therefore, the Harderian gland can produce toxins, occurring in the secretion at least in low levels. The function of some snake's toxins has been demonstrated as effective antimicrobial compounds [527]. Indeed, a SVMP from the Chinese viper from *Agkistrodon halys*, resulted more active than conventional drugs against some multi-drug resistant human pathogens

[528]. Thus, even the amount of toxins is not enough to subdue preys, they could be playing a protective role against pathogens in the orbital space and mucosal membranes. These evidences suggested that toxins are deeply involved in the Harderian gland function, becoming more intriguing its role.

5.5 Conclusions

Here, we used the NGS approach combined with bioinformatic tools to effectively characterize the transcriptome of the Harderian glands and deduce its possible biological function across three colubrid snakes (*sensu stricto*) from Cuba: *C. andreae*, *C. cantherigerus* and *T. variabilis*. Most of representative genes found in the transcriptome corresponded to lipocalin-like, bactericidal proteins, cystatin-like and toxins. Although, transcripts associated with housekeeping genes like mitochondrial, ribosomal and poorly characterized proteins also occurred within the 25 MEC from each of the colubrid species.

In general, some common proteins related to ion channels and binding proteins like actin were identified within the 1000 MEC as well as enzymes like proteases, phospholipases and oxidoreductases. Nonetheless, beyond the mentioned components that are almost ubiquitous, significant expression of other transcripts should be highlighted, such as lipocalin/lipocalin-like toxins, bactericidal/permeability-increasing protein-like and even toxins. Our results suggest that the function of the Harderian gland may be greatly involved in transport/binding and defence. The fact that the GO term “Binding” resulted as the most significant within the category “Molecular Function” reinforces the previous findings that assign Harderian gland with vomerolfaction, seemingly related to the transport of pheromones. Nonetheless, our results suggest it may be also associated with the production of toxins, which probably provide protection against pathogens acting in the orbital space and mucosal membranes or even to subdue preys. However, to elucidate the Harderian gland function, further proteomics approaches, namely high-throughput analyses of the whole gland and its secretion would be needed.

5.6 Supplementary Materials

Supplementary materials of the Chapter 5 are available online at the following link: <https://goo.gl/WnANkU>

CHAPTER 6

General Discussion

General Discussion

The research performed in this thesis included the exploration of toxins repertoire of two main groups of venomous animals. For this purpose, general characterization of the extract and semi-purified fraction of two unexplored species of cnidarians from Cuba and Portugal was performed. In addition, transcriptome analyses of the Harderian gland from three colubrids species of Cuba were done. The obtained results and findings will contribute to the knowledge about toxins distribution, protein/toxins encoding-genes and, eventually will assist in providing an evolutionary overview of such species and its products. Altogether, the different methodologies applied comprising both classical and high-throughput approaches have provided relevant empiric insights for toxins discovery.

The Phylum Cnidarian have been considered as the largest venomous group of extant species and the distribution of peptide toxins seems to be ubiquitous among its members. Besides, the fact that most cnidarians species remain unexplored justifies that three chapters of this thesis were dedicated to this Phylum. The **Chapter 2**, **Chapter 3** and **Chapter 4**, were addressed to explore the diversity of peptides and toxins, and even its toxicological properties.

In **Chapter 2** the zoanthid *Zoanthus sociatus* (Zoanthidea, Hexacorallia, Anthozoa, Cnidaria) sampled in Cuba was studied. The *Z. sociatus* belongs to Zoanthiniaria, which is an order of the Subclass Hexacorallia, Class Anthozoa, and unlike sea anemone (order Actiniaria), its toxins diversity remains practically unknown [14, 176]. After semi purification of the whole body-extract, a low molecular weight fraction named ZsG50-III was obtained. The chromatogram of the extract was congruent with previously profiles from cnidarians, obtained in a similar system [114, 176]. Moreover, the molecular weight composition of the fraction revealed by MALDI-TOF/TOF MS, showed m/z values ranging from 700 to 6000 Da. However, most of the m/z values and the most intense signals corresponded to m/z less than 1000 Da. This result contrasts with what is expected for the molecular weight of the sea anemone toxins ranging mostly between 2 and 10 KDa [14, 22, 27, 176, 529].

Moreover, a preliminary attempt was made to identify by MS/MS some of the major components of the sample. Given the low level of purity it was not possible to reliably identify any compound by this technique. However, based on the clustering of components in the molecular weight range less than 1000 and the fragmentation pattern, it was possible to infer that the fraction contains a mixture of low molecular weight peptidic and non-peptidic components.

The resulted fraction named ZsG50-III was assayed in mice, resulting lethal in a dose-depending manner. The percentage of lethality *versus* the dose plot fitted to a dose-response sigmoid curve (Figure 2.2A). This result means that the components of the fraction

at low concentrations cause a little visible effect, but once reaching a threshold, a small increasing of the dose trigger the toxic effects observed and eventually death. During the toxicological test it was observed a decreasing of exploratory activity and a disordered breathing 10 minutes after injection in lower doses. Higher doses provoked spasms, palpitations, convulsions and dead in less than a minute after injection. Moreover, fecal and urinary incontinency were also detected in the two higher doses. These effects were observed before sudden dead, which was preceded by dyspnea and reduced motile activity, suggesting cardiac arrest as the main cause of lethality. Certain drugs have been reported to cause respiratory and cardiovascular complications before cardiac arrest in mammals [18], including some cnidarian toxins [19].

In addition, a non-lethal dose (600 µg/kg) of the fraction ZsG50-III decreased the time that produced the death by a lethal dose of KCl (Figure 2.3). The significant decreasing in death time when combining a sublethal dose of the fraction with a cardiotoxic agent such as KCl, also supports that the mechanism of action of the fraction enhances the lethal effect of such agent. Interestingly, it was previously reported that the crude extract of *Z. sociatus* contains low molecular weight compounds that inhibit Ca^{2+} influx to pancreatic beta cells [114]. In addition, it is well known that overdose of Ca^{2+} channel blockers cause cardiovascular failure [327, 328]. Altogether, these evidences suggest that lethal effects are likely mediated by a cardiotoxicity process.

Thus, the rapid onset of the toxic effects once the dose reaches a threshold may be associated with the presence of low molecular weight compounds and their rapid absorption. Once reached the threshold the effects are produced in a short time range that ends in death or the recovery of the animal in the case of higher or sub-lethal doses, respectively. This suggests that the compound is rapidly absorbed, then binding to the receptor provoking reversible physiological alteration. However, a limitation of this study consisted in the insufficient evidence to allow the identification of the compounds responsible by such effects. In this sense, further analytical steps were needed.

Thus, in **Chapter 3** the purification of the low molecular weight fraction previously studied in **Chapter 2** was performed. To figure out the fraction compositions some purification steps comprised the separation of early-eluting compounds by RP-HPLC analyses followed by MALDI-TOF/TOF approaches. The mass spectrometry of some RP-HPLC revealed scarce peaks that resemble cyclic-peptides fragmentation [530] rather than sea anemone-like toxins [158, 176]. Nonetheless, the absence of sea anemone toxins does not mean its total absence, since only early eluting peaks were analyzed. Besides, the detection can be affected if the toxins are in a low concentration. Moreover, some signals in the range of sea anemones neurotoxin with low intensity were previously detected using linear mode in mass spectrometry analyses (**Chapter 2**). However, those signals were not

detected through positive ion reflector mode by MALDI-TOF/TOF. This effect could happen when the reflector mode is used to gain in resolution at the expense of sensitivity [176]. Although the concentration of toxins may be the main reason, the use of MALDI-TOF/TOF, specifically in reflector mode could be a limitation [531-534]. In contrast to our results, the spectra of a fraction from *Z. sociatus* obtained in similar conditions but using ESI, showed most intense signal ranging from 3 to 7 KDa [114]. Since detection is based on mass-to-charge (m/z) ratios, multiply charged ions are a desirable feature conferring an apparent low-molecular m/z to those high molecular weight proteins [535]. Thus, ESI must be an advantage in this case, because higher-molecular-weight compounds having masses outside the range of the mass spectrometer could be analyzed [535].

It is also noteworthy that while this species belongs to the Class Anthozoa, it is not exactly a member of the family Actiniaria where the marine sea anemones belong [157, 313, 397]. However, the presence of nematocysts in all cnidarians is a fact [156, 313]. Despite the production of toxins in this group seem to be universal [14], probably species belonging to this order have a different strategy. Its toxins production rather than be typical of sea anemones could be more like jellyfish toxins, showing high molecular weight. This assumption is also supported by the lack of sea anemones toxins, since the toxins described from this order are certainly non-proteinaceous. Indeed, Zoanthidea are a rich source of some potent non-peptide toxins, such as palytoxin and its analogues [100, 332, 333]. The presence of sea anemones-like toxins in the order Zonathidea is still unclear. Although, at the transcriptional level it was detected a sea anemone-like toxin called ShK/Aurelin-like that was toxic to zebrafish embryos [335]. However, some fraction from *Z. sociatus* not studied in this work could contain sea anemones toxins, such as those showing less absorbance in the chromatogram [176]. In this sense, future studies are needed to better determine the absence of sea anemones toxins in zoanthids.

Unexpectedly, most of the resulting masses and fragmentation pattern were related to cyanotoxins. Indeed, microcystin-LR was successfully identified. Despite being not clear if the MC-LR and other cyanotoxins presented in the fraction were responsible for the toxic effects previously reported (in **Chapter 2**), microcystins are highly toxic in mammals, specifically in mice [388]. The lethal dose MC-LR and its congeners administered intraperitoneally is 50 $\mu\text{g/kg}$, which classify it as highly toxic [389]. However, this study lack of enough information about the time in which such cyanotoxins produce the lethal effects. In addition, the toxic effect reported for MC-LR are hepatotoxic rather than the cardiotoxic effect proposed in **Chapter 2**. However, it is known that MC-LR can decrease the opening probability of large conductance K^+ (BK) channels via inhibition of endogenous phosphatases [536]. Besides, these potassium channels are voltage-dependent but can also be activated by a rising of intracellular Ca^{2+} [537, 538]. MC-LR has been pointed out as

causing potential neurotoxicity, since can also regulates Ca^{2+} activated phosphatase calcineurin (CaN) [539]. It is noteworthy, that a low molecular fraction from *Z. sociatus* obtained by similar methodologies produced Ca^{2+} influx blockade in isolated rat β cells [114]. Therefore, the presence of MC-LR in ZsG50-III could be involved in cardiotoxic effects trough Ca^{2+} and K^{+} channels modulation. However, the lack of enough amount of fraction was a limitation to addresses further analyses, hindering the identification of more cyanotoxins and its assessment *in vivo* or in toxicological tests.

In **Chapter 2** and **Chapter 3** both analytical and toxicological assays of a low molecular weight fraction from a cnidarian species were performed, given insight about the composition and toxicological properties of such fraction. The previous methodologies comprised some step from species identification, sampling, extract preparation, purification, proteomic analyses and toxicological tests. These approaches are useful but are time-consuming and request large amounts of biomass, which represented a limiting factor in **Chapters 2** and **3**. Unlike the mentioned approaches, shotgun proteomics needs less amounts of starting material to identify the components of a complex sample, but does not necessarily requires the previous purification steps [212]. Thus, in **Chapter 4** it was explored the benefits of high-throughput proteomics to profile the whole-body extract of the unexplored sea anemone *Bunodactis verrucosa* from Portugal. For this purpose, LC-MS/MS was carried out by nano-LC coupled to a hybrid Ion trap mass spectrometer, commonly known as Orbitrap (LTQ Orbitrap Velos Pro –ETD, Thermo Scientific, Waltham, MA, USA). In addition, for comparative aims also gel-based analyses were performed, combining two-dimensional gel electrophoresis (2DE) with MALDI-TOF/TOF.

Contrasting with previous spectrometry analyses (**Chapter 2** and **3**), in **Chapter 4** it was possible to identified several proteins (Figure 4.1; Table 4.1), including toxins (Table 4.2). Considering the number of protein identified versus time and amount of sample needed, shotgun proteomics resulted more suitable than gel-based analyses. It is known that LC-MS/MS is likely the most powerful techniques in the field of proteomics, since it yield the identification of many proteins out of the complex protein mixtures [155]. Similar techniques that employed nano liquid chromatography (nanoLC-MS/MS) were previously used with success in the study of cnidarians toxins [155, 167]. However, LC-MS/MS was applied by the first time in *B. verrucosa* providing a complete map of its proteome. Hence, most of the identified proteins represent the first report for this species.

The eight proteins from two-dimensional electrophoresis combined with MALDI-TOF/TOF were related to Glycolytic pathway, antioxidants activities and RNA degradation. This result was similar to that obtained with the Blast2Go after analyzing the data set from shotgun proteomics. Most represented pathways from the 412 proteins identified in such analyses, were Purine, Thiamine metabolism, Biosynthesis of antibiotics,

Glycolysis/Gluconeogenesis and Carbon fixation in photosynthetic organisms. The Purine metabolism pathway is closely related to the metabolism of nucleotides [540], since purine constitutes subunits of nucleic acids and precursors for the synthesis of nucleotide cofactors, while Thiamine metabolism pathway is fundamental in the metabolism of carbohydrates [430]. In summary, it can be observed that some proteins obtained by both methods of gel-based and gel-free, are involved in the same pathways. Some peptides and proteins involved in such roles were detected by both methods. It is also noteworthy that the most represented pathways are of great importance in the metabolism of the species.

Briefly, the Purine metabolism pathway is closely related to the metabolism of nucleotides [540], while the Thiamine metabolism pathway is fundamental for the metabolism of carbohydrates [430]. Besides, the last route mentioned in addition to Glycolysis/Gluconeogenesis and Carbon fixation in photosynthetic organisms' pathways are obviously of great importance for obtaining energy. The presences of symbionts such as zooxanthellae in sea anemones [541, 542], can explain the occurrence of the Carbon fixation in the photosynthetic organisms' pathway. In addition, one of the most significant pathways was the Biosynthesis of antibiotics. The presence of proteins associated with antibiotic function, must play a significant role in the defense against pathogens. Interestingly, most of the protein identified in this pathway belongs to the larval stage of *N. vectensis*, suggesting that antibiotics are expressed more in the early vulnerable stages. This ratifies sea anemones as a potential source of antibiotics [431, 433, 543], which can be used against resistant germs to current treatments [544].

Among all peptides generated, 63 sequences matched for 58 potential toxins (Supplementary materials **Chapter 4**, but only five toxins were identified with two peptides (Table 4.3). In the identification of the toxins, we were more conservative, thus only those protein reconstructed with more than two peptides (unique+razor) were accepted as potential toxins and discussed here. In addition, to increase confidence, we have verified that these peptides were not redundant to those proteins identified against previous cnidarians custom databases. Given that the database used for the identification of toxins was different from the one applied in the previous analysis, this approach avoids accepting a peptide that may belong to another protein in which its sequence origin has a better explanation than the suggested toxin (Occam's razor or Occam's principle). In resume, of the five toxin, two are metalloproteinases, one zinc metalloproteinase/disintegrin (VM2M2_DEIAC) of the snake *Deinagkistrodon acutus* [436] and Neprilysin-1 (NEP_TRILK) from brush-footed trapdoor spider *Trittame loki* [437]. Other was a phospholipase A2 (PLA₂) called vurt toxin (PA2B_VIPRE) from steppe viper *Vipera renardi* [545] and two putative neurotoxins; the alpha-latroinsectotoxin-Lt1a (LITA_LATTR) from the black widow spider

Latrodectus mactans tredecimguttatus [546] and SE-cephalotoxin (CTX_SEPES) from the salivary gland of cuttlefish *Sepia esculenta* [104].

Knowing in advance the ecological aspects of *B. verrucosa*, like its putative preys including mussels, crabs and gobies, it was possible to associate such toxins with its likely role in feeding and defense. Specifically, it seems that mussels are the main food resource of this anemone, since we observe this species regurgitating them during sampling. Furthermore, mussels are abundant in these intertidal pools where this anemone inhabits. This is not an isolated case, since other anemones like *Anthopleura elegantissima* and *Anthopleura xanthogrammica* in this ecosystem have mussels as their main resource of nourishment [449, 450]. The presence of these toxins is sufficient to explain the mechanism of feeding and defense of the species under study. However, it should be noted that other hydrolytic enzymes like hyaluronidases and proteases found in *B. verrucosa* may be secreted into the gastrovascular cavity acting together with the five toxins identified. Also, the presence of other toxins should not be discharged, instead addressed further in future explorations.

Briefly, once a mussel is captured it will tightly close its valves. Then, enzymes like zinc metalloproteinase/disintegrin (VM2M2_DEIAC) and Neprilysin-1 (NEP_TRILK) could degrade the tissues that seals the shell, probably a dorsal elastic proteinaceous-ligament extending for the length of the hinge [454]. The tissues degradation by metalloproteinases can facilitate the diffusion of neurotoxins inside the prey. Then, neurotoxins like SE-cephalotoxin and PLA₂ vurt toxin could act on the adductor muscle, causing the loss of its function and leading to the valve opening. SE-cephalotoxin can produce flaccid paralysis increasing the valves gape aperture, while PLA₂ vurt toxin could block the adductor muscles. These inferences were based on the properties previously described for each toxin (see **Chapter 4**). A similar mechanism involving the metalloproteinases that facilitate the neurotoxins diffusion was previously proposed in the spider *Trittame loki* [437].

Nonetheless, this methodology requires a reference genome to be faithful. In addition, without previous purification any fraction is pooled, which precludes further toxicological and pharmacological characterization of the proteins or toxins identified. For this reason, even though databases of phylogenetically close species have been used, a genome or transcriptome of the species under study would be a better option.

In this sense, the need for methodologies capable of providing wide information about proteins and its encoding genes serving also as reference databases for shotgun proteomics analyses, led us to explore transcriptomic approaches. The mentioned elements combined with lack of high-throughput screening addressed to snakes from Cuba, drove us to profile the transcriptome of the Harderian glands from Cuban colubrids. In **Chapter 5** it was assessed the whole transcriptome of the Harderian gland of three colubrids from Cuba.

These glands were selected because the information regarding its function is scarce and have also been recognized as the most enigmatic organ from vertebrates, which its function remaining unknown.

In summary, a total of 54,593,038 reads from *Caraiba andreae*, 50,762,546 from *Cubophis cantherigerus* and 56,115,370 from *Tretanorhinus variabilis* were obtained using Illumina HiSeq2000 100 bp (base-pairs) paired-end (PE). The large amount of reads yielded with a high accuracy provided by Illumina are congruent with those commonly used by this platform, which explains their preference by researchers [244]. The high number of reads obtained represents also an advantages for *de novo* assembly [244]. In this study, 63,333 contigs resulted from *de novo* assembly in *C. andreae*, while 27,666 assembled contigs were obtained from *C. cantherigerus* and 62,324 contigs from *T. variabilis*. The average contigs length (N50) were 1092, 1602 and 1337 bp for *C. andreae*, *C. cantherigerus* and *T. variabilis*, respectively. Despite some metrics largely used in genomic analysis seems to be unlikely to evaluate the quality of a transcriptome assembly [492], such statistics (see Table 5.1) constitute an evidence of the faithful protocol applied which included, sample preparation, library synthesis, platform selected and *de novo* assembly strategy.

Then the annotation process and expression analyses revealed as highly expressed genes related to uncharacterized proteins, gene encoding mitochondrial proteins, ribosomal proteins, lipocalin-like proteins, bactericidal proteins, cystatin-like proteins and some toxins. Some of these genes could be considered as housekeeping genes, since they are associated with basic life-sustaining processes, like basal metabolism, DNA and protein synthesis. On the contrary, genes related to toxins like lipocalin-like proteins, cystatin-like proteins and toxins suggest that the Harderian gland should be greatly involved in transport/binding and defense. This assumption was reinforced because the GO term “Binding” resulted as the most significant within the category “Molecular Function”. Furthermore, deeper analysis of the Harderian’s gland toxin composition revealed most classes of the snakes’ toxins expressed into the transcriptome. Thus, protection against pathogens could be one of the most important functions of the Harderian gland, considering that bactericidal proteins were also found highly expressed with the toxins and the other components of venom. The secretion containing such compounds and toxins could provide a mechanism of defense against germs in the orbital space and mucosal membranes. However, as the results obtained were just at the transcriptional level, it is not yet clear if the secretion of the gland contains toxins playing a role in subduing the preys.

CHAPTER 7

Conclusions and Future Perspectives

Concluding Remarks

In this thesis, different methodologies were explored to characterize the toxicological properties and toxins production of two main groups of venomous animals: cnidarians and snakes. The methodologies comprised mainly proteomics and transcriptomic approaches, including some protocols from classical methods based on extract purification, gel-free and gel-based proteomic methods and toxicological tests in mice.

In **Chapter 2** and **3** it was studied a low molecular weight fraction from *Z. sociatus*. MALDI TOF/TOF analyses revealing a composition of mostly masses between 700-1000 Da. Among such masses no evidence of sea anemones-like toxins was found. On the contrary, the most intense m/z signals of the fraction are related to microcystins. In this sense, microcystin MC-LR was identified by tandem mass spectrometry analyses. The origin of such microcystins in *Z. sociatus* is unclear but it is likely from external source. Since MC-LR is highly toxic to mammals, specifically to mice, the possible implication of such microcystins in the lethal effects reported from the studied fraction ZsG50-III was discussed.

However, the last assumption needs further evaluation, which would require more biomass. If enough biomass would be available, then it would be possible to purify extracts and characterize the bioactivity of each fraction as bioassay-guided. Obtaining a pure fraction before conducted to spectrometry analyses would be preferred. If the sample is not pure at all and the components have low molecular weight (less than 1000 Da), then MALDI would be more appropriate. If the sample shows higher molecular weight compounds, then ESI would be more suitable.

In the **Chapter 4**, the whole proteome of the sea anemone *B. verrucosa* was characterized by the first time. In general, shotgun proteomics resulted more suitable yielding more peptide sequences than gel-based analyses. Altogether, a total of 413 proteins were identified by shotgun proteomics gel-based and gel-free approaches. Some of the proteins identified corresponded to pathways associated with Purine, Thiamine metabolism, Biosynthesis of antibiotics, Glycolysis/Gluconeogenesis and Carbon fixation in photosynthetic organisms. Some of the proteins identified in these pathways were detected in both gel-based and gel-free methods, which provided a complementary inference. However, the higher number of enzymes involved in the KEEG analyses corresponded to the Biosynthesis of antibiotics pathway. This finding confirms the importance of engineered biological compounds used in defense.

On the other hand, some potential toxins like metalloproteinases were identified, as well as neurotoxin like SE-cephalotoxin. The combination of proteomic evidences and the ecology of the species, shed light about its strategy to subdue preys like mussels. In this sense, the toxins seemingly act synergically. Metalloproteinase and the other hydrolytic

enzymes detected may produce a degradation of the tissues, facilitating the neurotoxins access to the target, which produces muscle paralysis. These effects, should release the valves allowing the sea anemone to obtain nourishment.

In the **Chapter 5** the NGS approach combined with bioinformatic tools resulted effectively to profile the transcriptome of the Harderian glands of three Cuban colubrids: *Caraiba andreae*, *Cubophis cantherigerus* and *Tretanorhinus variabilis*. Some genes were found highly expressed and associated with lipocalin-like, bactericidal proteins, cystatin-like and toxins. Moreover, housekeeping genes like mitochondrial, ribosomal, ion channel, binding and poorly characterized proteins occurred within the most expressed transcripts. In addition, the functional annotation obtained by GO analyses, revealed GO term “Binding” as the most significant within the category “Molecular Function”. The combination of our results reinforces the previous assumption about the Harderian gland involvement in transport/binding. However, the presence of bactericidal/permeability-increasing protein-like and toxins among the most expressed genes, suggest a new function for this gland likely related to the protection against pathogens in the orbital space and the mucosal membranes. The putative role of the gland in toxins production can also assist the snake to subdue preys, but this function need to be further evaluated.

Future Perspectives

As a recommendation for future works related to **Chapter 2 and 3**, the specimens must be carefully carry out to the lab and frozen until use. After extract preparations, Sephadex and cation-exchange must be used, and then two different systems need to be applied for RP-HPLC analyses. One system can be similar like the one reported here using methanol as mobile phase and other in acetonitrile gradient elution system. These steps can be combined with bioassays-guided chromatographic purifications, toxicological and pharmacological tests (in an *in vivo* and *in vitro* model). Toxicological assays must be repeated in mice with the fraction pooled. In the case of pure fractions or toxins, heterologous expression system in *Xenopus* oocytes for studying ion channels modulation could be used. The isolation of nematocysts to analyze its composition with nanoLC-MS/MS system combined with its transcriptome it is also recommended as future perspectives. If there is not enough biomass, the combination of shotgun proteomic and transcriptomics would be the best option. This perspective will help to unravel the composition of *Z. sociatus* toxins and the distribution of sea anemones toxins in Zoanthids, and in the Phylum Cnidaria.

The future perspective for **Chapter 4** is to perform both the whole and isolated-nematocysts transcriptome to increase the number and accuracy of the protein/toxin identified. In addition, to corroborate the effects of the toxins, mussels can be used as *in vivo*

model. Moreover, to unravel the mechanism of action proposed, similar bioassays-guided chromatographic purifications steps can be used as recommended in **Chapter 2 and 3**.

On the other hand, as the results obtained in **Chapter 5** were just at the transcriptional level, the function of the Harderian gland is still intriguing. Thus, proteomics approaches addressing gland secretion would be welcome in the near future. Therefore, transcriptomic analyses combined with shotgun proteomics approaches are needed for better understanding the role of such gland. As biomass of the glands and its secretions is limited, nanoLC-MS/MS system combined with transcriptomic analyses constitute the best choice.

General Conclusions

Overall, the results obtained in this thesis led us to conclude that the combination of high-throughput approaches with classical bioassays-guided chromatographic purifications provided an integrated view for protein characterization. High-throughput methodologies like shotgun proteomics combined with transcriptomics constitutes powerful tools, allowing the detection of proteins in large-scale, with their respective encoding-genes. The resulting information from the mentioned approaches also provided enough information to unravel a tissue role. Although, such information can also be used as platform for evolutionary analyses to unravel protein/gene function. However, even when such methodology shed light from a complex sample, the classical methods of protein characterization are still needed. The bioassays-guided chromatographic purifications allow evaluating the toxicological and pharmacological properties of the fractions pooled, which constitutes the starting point for further biomedical applications. In conclusion, it is strongly recommended to integrate as much as possible the methodologies applied in this work. Integrated omics perspectives will allow unraveling the tissue proteome and its function, complete isolation and characterization of specific protein or toxins. Moreover, the resulting data from these approaches provides a wide overview of protein/toxins-encoding genes, which reveal insights of its evolutionary history and the genetics mechanism involved in the biological evolution.

CHAPTER 8

References

References

1. Ménez A, Stöcklin R, Mebs D: **'Venomics' or: the venomous systems genome project.** In.: Pergamon; 2006.
2. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annual review of genomics and human genetics* 2009, **10**:483-511.
3. Fox JW, Serrano SM: **Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures.** *Proteomics* 2008, **8**(4):909-920.
4. Fry BG: **From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins.** *Genome research* 2005, **15**(3):403-420.
5. Inceoglu B, Lango J, Jing J, Chen L, Doymaz F, Pessah IN, Hammock BD: **One scorpion, two venoms: pre venom of *Parabuthus transvaalicus* acts as an alternative type of venom with distinct mechanism of action.** *Proceedings of the National Academy of Sciences* 2003, **100**(3):922-927.
6. Olivera BM, Teichert RW: **Diversity of the neurotoxic *Conus* peptides.** *Molecular Interventions* 2007, **7**(5):251.
7. Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF: **Were arachnids the first to use combinatorial peptide libraries?** *Peptides* 2005, **26**(1):131-139.
8. Mebs D: **Venomous and poisonous animals: a handbook for biologists, toxicologists and toxinologists, physicians and pharmacists:** Medpharm; 2002.
9. Harvey AL: **Presynaptic effects of toxins.** *International review of neurobiology* 1990, **32**:201-239.
10. Harvey A, Bradley K, Cochran S, Rowan E, Pratt J, Quillfeldt J, Jerusalinsky D: **What can toxins tell us for drug discovery?** *Toxicon* 1998, **36**(11):1635-1640.
11. Hodgson WC, Isbister GK: **The application of toxins and venoms to cardiovascular drug discovery.** *Current opinion in pharmacology* 2009, **9**(2):173-176.
12. Karalliedde L: **Animal toxins.** *British Journal of Anaesthesia* 1995, **74**(3):319-327.
13. Froy O, Sagiv T, Poreh M, Urbach D, Zilberberg N, Gurevitz M: **Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels.** *Journal of molecular evolution* 1999, **48**(2):187-196.
14. Honma T, Shiomi K: **Peptide toxins in sea anemones: structural and functional aspects.** *Marine biotechnology* 2006, **8**(1):1-10.
15. Lynch VJ: **Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A 2 genes.** *BMC evolutionary biology* 2007, **7**(1):2.
16. Kordiš D, Gubenšek F: **Adaptive evolution of animal toxin multigene families.** *Gene* 2000, **261**(1):43-52.
17. Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney A: **Molecular evolution and phylogeny of elapid snake venom three-finger toxins.** *Journal of molecular evolution* 2003, **57**(1):110-129.
18. Bosmans F, Tytgat J: **Sea anemone venom as a source of insecticidal peptides acting on voltage-gated Na⁺ channels.** *Toxicon* 2007, **49**(4):550-560.
19. Dauplais M, Lecoq A, Song J, Cotton J, Jamin N, Gilquin B, Roumestand C, Vita C, de Medeiros CL, Rowan EG: **On the convergent evolution of animal toxins conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures.** *Journal of Biological Chemistry* 1997, **272**(7):4302-4309.
20. Gasparini S, Gilquin B, Ménez A: **Comparison of sea anemone and scorpion toxins binding to Kv1 channels: an example of convergent evolution.** *Toxicon* 2004, **43**(8):901-908.

21. Ruoslahti E: **RGD and other recognition sequences for integrins.** *Annual review of cell and developmental biology* 1996, **12**(1):697-715.
22. Frazão B, Vasconcelos V, Antunes A: **Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview.** *Marine drugs* 2012, **10**(8):1812-1851.
23. Duda TF, Palumbi SR: **Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*.** *Proceedings of the National Academy of Sciences* 1999, **96**(12):6820-6823.
24. Nakashima K-i, Ogawa T, Oda N, Hattori M, Sakaki Y, Kihara H, Ohno M: **Accelerated evolution of *Trimeresurus flavoviridis* venom gland phospholipase A2 isozymes.** *Proceedings of the National Academy of Sciences* 1993, **90**(13):5964-5968.
25. Zhu S, Bosmans F, Tytgat J: **Adaptive evolution of scorpion sodium channel toxins.** *Journal of molecular evolution* 2004, **58**(2):145-153.
26. Moran Y, Weinberger H, Sullivan JC, Reitzel AM, Finnerty JR, Gurevitz M: **Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the *Nematostella vectensis* genome.** *Molecular biology and evolution* 2008, **25**(4):737-747.
27. Norton RS: **Structures of sea anemone toxins.** *Toxicon* 2009, **54**(8):1075-1088.
28. Moran Y, Gordon D, Gurevitz M: **Sea anemone toxins affecting voltage-gated sodium channels—molecular and evolutionary features.** *Toxicon* 2009, **54**(8):1089-1101.
29. Gordon D, Karbat I, Ilan N, Cohen L, Kahn R, Gilles N, Dong K, Stühmer W, Tytgat J, Gurevitz M: **The differential preference of scorpion α -toxins for insect or mammalian sodium channels: implications for improved insect control.** *Toxicon* 2007, **49**(4):452-472.
30. Nicholson GM: **Insect-selective spider toxins targeting voltage-gated sodium channels.** *Toxicon* 2007, **49**(4):490-512.
31. Ligabue-Braun R, Verli H, Carlini CR: **Venomous mammals: a review.** *Toxicon* 2012, **59**(7):680-695.
32. Schmidt JO: **Venom and the good life in tarantula hawks (Hymenoptera: Pompilidae): how to eat, not be eaten, and live long.** *Journal of the Kansas Entomological Society* 2004, **77**(4):402-413.
33. Petrunkevitch A: **Tarantula versus tarantula-hawk: A study in instinct.** *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* 1926, **45**(2):367-397.
34. Wang S-Y, Wang GK: **Voltage-gated sodium channels as primary targets of diverse lipid-soluble neurotoxins.** *Cellular signalling* 2003, **15**(2):151-159.
35. Cestèle S, Catterall WA: **Molecular mechanisms of neurotoxin action on voltage-gated sodium channels.** *Biochimie* 2000, **82**(9):883-892.
36. Hille B: **Ion channels of excitable membranes**, vol. 507: Sinauer Sunderland, MA; 2001.
37. Dunne M, Petersen O: **Potassium selective ion channels in insulin-secreting cells: physiology, pharmacology and their role in stimulus-secretion coupling.** *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 1991, **1071**(1):67-82.
38. Dhaka A, Viswanath V, Patapoutian A: **Trp ion channels and temperature sensation.** *Annu Rev Neurosci* 2006, **29**:135-161.
39. Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ: **Principles of neural science**, vol. 4: McGraw-hill New York; 2000.
40. Kuffler SW, Nicholls JG, Martin A: **A cellular approach to the function of the nervous system.** *Massachusetts: Sinauer Assoc* 1976.
41. Moran MM, Xu H, Clapham DE: **TRP ion channels in the nervous system.** *Current opinion in neurobiology* 2004, **14**(3):362-369.

42. Caldwell JH, Schaller KL, Lasher RS, Peles E, Levinson SR: **Sodium channel Nav1.6 is localized at nodes of Ranvier, dendrites, and synapses.** *Proceedings of the National Academy of Sciences* 2000, **97**(10):5616-5620.
43. Liebeskind BJ, Hillis DM, Zakon HH: **Evolution of sodium channels predates the origin of nervous systems in animals.** *Proceedings of the National Academy of Sciences* 2011, **108**(22):9154-9159.
44. Liu D-W, Gintant GA, Antzelevitch C: **Ionic bases for electrophysiological distinctions among epicardial, midmyocardial, and endocardial myocytes from the free wall of the canine left ventricle.** *Circulation Research* 1993, **72**(3):671-687.
45. Salinas EM, Cebada J, Valdés A, Garateix A, Aneiros A, Alvarez JL: **Effects of a toxin from the mucus of the Caribbean sea anemone (*Bunodosoma granulifera*) on the ionic currents of single ventricular mammalian cardiomyocytes.** *Toxicon* 1997, **35**(12):1699-1709.
46. Campbell C: **The effects of snake venoms and their neurotoxins on the nervous system of man and animals.** *Contemporary neurology series* 1974, **12**:259-293.
47. Beeler GW, Reuter H: **Reconstruction of the action potential of ventricular myocardial fibres.** *The Journal of physiology* 1977, **268**(1):177.
48. Frank HY, Yarov-Yarovoy V, Gutman GA, Catterall WA: **Overview of molecular relationships in the voltage-gated ion channel superfamily.** *Pharmacological reviews* 2005, **57**(4):387-395.
49. Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P: **Essential cell biology:** Garland Science; 2013.
50. Possani LD, Merino E, Corona M, Bolivar F, Becerril B: **Peptides and genes coding for scorpion toxins that affect ion-channels.** *Biochimie* 2000, **82**(9):861-868.
51. Bähring R, Covarrubias M: **Mechanisms of closed-state inactivation in voltage-gated ion channels.** *The Journal of physiology* 2011, **589**(3):461-479.
52. Harvey A, Karlsson E: **Dendrotoxin from the venom of the green mamba, *Dendroaspis angusticeps*.** *Naunyn-Schmiedeberg's archives of pharmacology* 1980, **312**(1):1-6.
53. Joubert F, Taljaard N: **The amino acid sequence of two proteinase inhibitor homologues from *Dendroaspis angusticeps* venom.** *Hoppe Seylers Z Physiol Chem* 1980, **361**:661-674.
54. Harvey A, Robertson B: **Dendrotoxins: structure-activity relationships and effects on potassium ion channels.** *Current medicinal chemistry* 2004, **11**(23):3065-3072.
55. Benishin CG, Sorensen RG, Brown WE, Krueger BK, Blaustein MP: **Four polypeptide components of green mamba venom selectively block certain potassium channels in rat brain synaptosomes.** *Molecular pharmacology* 1988, **34**(2):152-159.
56. Marshall D, Harvey A: **Protease inhibitor homologues of dendrotoxin do not bind to dendrotoxin acceptors on synaptosomal membranes or facilitate neuromuscular transmission.** *Biological chemistry Hoppe-Seyler* 1992, **373**(8):707.
57. Harvey A, Karlsson E: **Protease inhibitor homologues from mamba venoms: facilitation of acetylcholine release and interactions with prejunctional blocking toxins.** *British journal of pharmacology* 1982, **77**(1):153-161.
58. Oliveira JS, Fuentes-Silva D, Zaharenko A: **Sea anemone peptides. Biological activities, structure-function relationships and phylogenetic aspects.** *Animal toxins: state of the art perspective in health and biotechnology 1st ed* Belo Horizonte: Editora UFMG 2009.
59. Catterall WA, Cestèle S, Yarov-Yarovoy V, Frank HY, Konoki K, Scheuer T: **Voltage-gated ion channels and gating modifier toxins.** *Toxicon* 2007, **49**(2):124-141.
60. Catterall WA, Beress L: **Sea anemone toxin and scorpion toxin share a common receptor site associated with the action potential sodium ionophore.** *Journal of Biological Chemistry* 1978, **253**(20):7393-7396.

61. Gordon D, Zlotkin E: **Binding of an α scorpion toxin to insect sodium channels is not dependent on membrane potential.** *FEBS letters* 1993, **315**(2):125-128.
62. Little MJ, Wilson H, Zappia C, Cestèle S, Tyler MI, Martin-Eauclaire M-F, Gordon D, Nicholson GM: **δ -Atracotoxins from Australian funnel-web spiders compete with scorpion α -toxin binding on both rat brain and insect sodium channels.** *FEBS letters* 1998, **439**(3):246-252.
63. Jover E, Martin-Moutot N, Couraud F, Rochat H: **Binding of scorpion toxins to rat brain synaptosomal fraction. Effects of membrane potential, ions, and other neurotoxins.** *Biochemistry* 1980, **19**(3):463-467.
64. Levitan IB, Kaczmarek LK: **The neuron: cell and molecular biology:** Oxford University Press, USA; 2015.
65. Diochot S, Baron A, Rash LD, Deval E, Escoubas P, Scarzello S, Salinas M, Lazdunski M: **A new sea anemone peptide, APETx2, inhibits ASIC3, a major acid-sensitive channel in sensory neurons.** *The EMBO journal* 2004, **23**(7):1516-1525.
66. Andreev YA, Kozlov SA, Koshelev SG, Ivanova EA, Monastyrnaya MM, Kozlovskaya EP, Grishin EV: **Analgesic compound from sea anemone *Heteractis crispa* is the first polypeptide inhibitor of vanilloid receptor 1 (TRPV1).** *Journal of Biological Chemistry* 2008, **283**(35):23914-23921.
67. Diochot S, Baron A, Salinas M, Douguet D, Scarzello S, Dabert-Gay A-S, Debayle D, Friend V, Alloui A, Lazdunski M: **Black mamba venom peptides target acid-sensing ion channels to abolish pain.** *Nature* 2012, **490**(7421):552-555.
68. Terlau H, Olivera BM: **Conus venoms: a rich source of novel ion channel-targeted peptides.** *Physiological reviews* 2004, **84**(1):41-68.
69. Escoubas P, King GF: **Venomomics as a drug discovery platform.** *Expert review of proteomics* 2009, **6**(3):221-224.
70. Miljanich G: **Ziconotide: neuronal calcium channel blocker for treating severe chronic pain.** *Current medicinal chemistry* 2004, **11**(23):3029-3040.
71. Nobile M, Magnelli V, Lagostena L, Mochca-Morales J, Possani L, Prestipino G: **The toxin helothermine affects potassium currents in newborn rat cerebellar granule cells.** *Journal of Membrane Biology* 1994, **139**(1):49-55.
72. Nobile M, Noceti F, Prestipino G, Possani LD: **Helothermine, a lizard venom toxin, inhibits calcium current in cerebellar granules.** *Experimental brain research* 1996, **110**(1):15-20.
73. Wang J, Shen B, Guo M, Lou X, Duan Y, Cheng XP, Teng M, Niu L, Liu Q, Huang Q: **Blocking Effect and Crystal Structure of Natrin Toxin, a Cysteine-Rich Secretory Protein from *Naja atra* Venom that Targets the BKCa Channel.** *Biochemistry* 2005, **44**(30):10145-10152.
74. Yamazaki Y, Koike H, Sugiyama Y, Motoyoshi K, Wada T, Hishinuma S, Mita M, Morita T: **Cloning and characterization of novel snake venom proteins that block smooth muscle contraction.** *The FEBS Journal* 2002, **269**(11):2708-2715.
75. Yamazaki Y, Brown RL, Morita T: **Purification and cloning of toxins from elapid venoms that target cyclic nucleotide-gated ion channels.** *Biochemistry* 2002, **41**(38):11331-11337.
76. Brown RL, Haley TL, West KA, Crabb JW: **Pseudechetoxin: a peptide blocker of cyclic nucleotide-gated ion channels.** *Proceedings of the National Academy of Sciences* 1999, **96**(2):754-759.
77. Brown RL, Lynch LL, Haley TL, Arsanjani R: **Pseudechetoxin Binds to the Pore Turret of Cyclic Nucleotide-gated Ion Channels.** *The Journal of general physiology* 2003, **122**(6):749-760.
78. Morrisette J, Krättschmar J, Haendler B, El-Hayek R, Mochca-Morales J, Martin BM, Patel JR, Moss RL, Schleuning W-D, Coronado R: **Primary structure and properties of helothermine, a peptide toxin that blocks ryanodine receptors.** *Biophysical journal* 1995, **68**(6):2280-2288.

79. Anderluh G, Sepčić K, Turk T, Maček P: **Cytolytic proteins from cnidarians-an overview.** *Acta Chimica Slovenica* 2011, **58**(4).
80. Alvarez C, Mancheño JM, Martínez D, Tejuca M, Pazos F, Lanio ME: **Sticholysins, two pore-forming toxins produced by the Caribbean Sea anemone *Stichodactyla helianthus*: their interaction with membranes.** *Toxicon* 2009, **54**(8):1135-1147.
81. Südhof TC: **α -Latrotoxin and its receptors: neurexins and CIRL/latrophilins.** *Annual review of neuroscience* 2001, **24**(1):933-962.
82. Vetter RS: **Myth: Idiopathic wounds are often due to brown recluse or other spider bites throughout the United States.** *Western Journal of Medicine* 2000, **173**(5):357.
83. Atkinson RK, Wright LG: **The involvement of collagenase in the necrosis induced by the bites of some spiders.** *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology* 1992, **102**(1):125-128.
84. Elgert K, Wright R, Campbell B, Barrett J: **Radiolabeling of polypeptide components of Missouri brown spider venom.** In: *Federation proceedings: 1973: Federation amer soc exp biol* 9650 rockville pike, bethesda, md 20814-3998; 1973: 223-&.
85. Finke J, Campbell B, Barrett J: **A diagnostic test for *Loxosceles reclusa* bites.** In: *Annual meeting of the American Society of Microbiology: 1973; 1973.*
86. Geren CR, Chan TK, Howell DE, Odell GV: **Isolation and characterization of toxins from brown recluse spider venom (*Loxosceles reclusa*).** *Archives of biochemistry and biophysics* 1976, **174**(1):90-99.
87. Auth PC: **Physician Assistant review:** Lippincott Williams & Wilkins; 2012.
88. Ramos-Cerrillo B, Olvera A, Odell GV, Zamudio F, Paniagua-Solís J, Alagón A, Stock RP: **Genetic and enzymatic characterization of sphingomyelinase D isoforms from the North American fiddleback spiders *Loxosceles boneti* and *Loxosceles reclusa*.** *Toxicon* 2004, **44**(5):507-514.
89. Forrester L, Barrett J, Campbell B: **Red blood cell lysis induced by the venom of the brown recluse spider: the role of sphingomyelinase D.** *Archives of biochemistry and biophysics* 1978, **187**(2):355-365.
90. Schaloske RH, Dennis EA: **The phospholipase A 2 superfamily and its group numbering system.** *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 2006, **1761**(11):1246-1259.
91. Nevalainen TJ: **Phospholipases A 2 in the genome of the sea anemone *Nematostella vectensis*.** *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 2008, **3**(3):226-233.
92. Kiyatkin N, Dulubova I, Grishin E: **Cloning and structural analysis of alpha-latroitoxin cDNA. Abundance of ankyrin-like repeats.** *European journal of biochemistry* 1993, **213**(1):121.
93. Kini RM: **Excitement ahead: structure, function and mechanism of snake venom phospholipase A 2 enzymes.** *Toxicon* 2003, **42**(8):827-840.
94. Pungerčar J, Križaj I: **Understanding the molecular mechanism underlying the presynaptic toxicity of secreted phospholipases A 2.** *Toxicon* 2007, **50**(7):871-892.
95. Šribar J, Oberčkal J, Križaj I: **Understanding the molecular mechanism underlying the presynaptic toxicity of secreted phospholipases A 2: An update.** *Toxicon* 2014, **89**:9-16.
96. Schweitz H, Bruhn T, Guillemare E, Moinier D, Lancelin J, Béress L, Lazdunski M: **Kalicludines and kaliseptine. Two different classes of sea anemone toxins for voltage sensitive K⁺ channels.** *The Journal of biological chemistry* 1995, **270**(42):25121.
97. Yuan C-H, He Q-Y, Peng K, Diao J-B, Jiang L-P, Tang X, Liang S-P: **Discovery of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas.** *PloS one* 2008, **3**(10):e3414.

98. Liang S: **An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [= *Ornithoctonus huwena* (Wang)]**. *Toxicon* 2004, **43**(5):575-585.
99. Antuch W, Berndt K, Chávez M, Delfín J, Wüthrich K: **The NMR solution structure of a Kunitz-type proteinase inhibitor from the sea anemone *Stichodactyla helianthus***. *European journal of biochemistry* 1993, **212**(3):675.
100. Beress L: **Biologically active compounds from coelenterates**. *Pure and Applied Chemistry* 1982, **54**(10):1981-1994.
101. Minagawa S, Ishida M, Shimakura K, Nagashima Y, Shiomi K: **Isolation and amino acid sequences of two Kunitz-type protease inhibitors from the sea anemone *Anthopleura aff. xanthogrammica***. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 1997, **118**(2):381-386.
102. Hisada M, Satake H, Masuda K, Aoyama M, Murata K, Shinada T, Iwashita T, Ohfune Y, Nakajima T: **Molecular components and toxicity of the venom of the solitary wasp, *Anoplius samariensis***. *Biochemical and biophysical research communications* 2005, **330**(4):1048-1054.
103. Schwartz EF, Diego-Garcia E, de la Vega RCR, Possani LD: **Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones)**. *BMC genomics* 2007, **8**(1):119.
104. Bayrhuber M, Vijayan V, Ferber M, Graf R, Korukottu J, Imperial J, Garrett JE, Olivera BM, Terlau H, Zweckstetter M: **Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family structural and functional characterization**. *Journal of Biological Chemistry* 2005, **280**(25):23766-23770.
105. Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SR, Vidal N, Poelmann RE, Norman JA: **Evolution of an arsenal structural and functional diversification of the venom system in the advanced snakes (Caenophidia)**. *Molecular & Cellular Proteomics* 2008, **7**(2):215-246.
106. Campbell CL, Wilson WC, Manninen K: **Characterization of differentially expressed midge genes in orbivirus vector populations**. In: *American journal of tropical medicine and hygiene: 2005: Amer soc trop med & hygiene 8000 westpark dr, ste 130, mclean, va 22101 USA; 2005: 49-49*.
107. Wei A, Alexander RS, Duke J, Ross H, Rosenfeld SA, Chang C-H: **Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa**. *Journal of molecular biology* 1998, **283**(1):147-154.
108. Krishnaswamy S: **The transition of prothrombin to thrombin**. *Journal of Thrombosis and Haemostasis* 2013, **11**(s1):265-276.
109. Diaz-Garcia CM, Sanchez-Soto C, Hiriart M: **Toxins that modulate ionic channels as tools for exploring insulin secretion**. *Cellular and molecular neurobiology* 2010, **30**(8):1275-1281.
110. Gonçalves AA, Toyama MH, Carneiro EM, Marangoni S, Arantes EC, Giglio JR, Boschero AC: **Participation of Na⁺ channels in the potentiation by *Tityus serrulatus* α -toxin TsTx-V of glucose-induced electrical activity and insulin secretion in rodent islet β -cells**. *Toxicon* 2003, **41**(8):1039-1045.
111. Aicardi G, Pollo A, Sher E, Carbone E: **Noradrenergic inhibition and voltage-dependent facilitation of ω -conotoxin-sensitive Ca channels in insulin-secreting RINm5F cells**. *FEBS letters* 1991, **281**(1-2):201-204.
112. Braun M, Ramracheya R, Bengtsson M, Zhang Q, Karanauskaite J, Partridge C, Johnson PR, Rorsman P: **Voltage-gated ion channels in human pancreatic β -cells: electrophysiological characterization and role in insulin secretion**. *Diabetes* 2008, **57**(6):1618-1628.
113. Herrington J: **Gating modifier peptides as probes of pancreatic β -cell physiology**. *Toxicon* 2007, **49**(2):231-238.
114. Diaz-Garcia CM, Sanchez-Soto C, Fuentes-Silva D, Leon-Pinzon C, Dominguez-Perez D, Varela C, Rodriguez-Romero A, Castañeda O, Hiriart M: **Low molecular weight compounds from *Zoanthus sociatus* impair insulin secretion via Ca²⁺**

- influx blockade and cause glucose intolerance in vivo.** *Toxicon* 2012, **59**(2):306-314.
115. Diaz-Garcia C, Fuentes-Silva D, Sanchez-Soto C, Domínguez-Pérez D, García-Delgado N, Varela C, Mendoza-Hernández G, Rodriguez-Romero A, Castaneda O, Hiriart M: **Toxins from *Physalia physalis* (Cnidaria) raise the intracellular Ca²⁺ of beta-cells and promote insulin secretion.** *Current medicinal chemistry* 2012, **19**(31):5414-5423.
 116. Nguyen TTN, Folch B, Létourneau M, Vaudry D, Truong NH, Doucet N, Chatenet D, Fournier A: **Cardiotoxin-I: An Unexpectedly Potent Insulinotropic Agent.** *ChemBioChem* 2012, **13**(12):1805-1812.
 117. Kim W, Egan JM: **The Role of Incretins in Glucose Homeostasis and Diabetes Treatment.** *Pharmacological reviews* 2008, **60**(4):470-512.
 118. Furman BL: **The development of Byetta (exenatide) from the venom of the Gila monster as an anti-diabetic agent.** *Toxicon* 2012, **59**(4):464-471.
 119. Wan Q-F, Dong Y, Yang H, Lou X, Ding J, Xu T: **Protein Kinase Activation Increases Insulin Secretion by Sensitizing the Secretory Machinery to Ca(2+).** *The Journal of General Physiology* 2004, **124**(6):653-662.
 120. Gao Z, Young RA, Trucco MM, Greene SR, Hewlett EL, Matschinsky FM, Wolf BA: **Protein kinase A translocation and insulin secretion in pancreatic beta-cells: studies with adenylate cyclase toxin from *Bordetella pertussis*.** *Biochemical Journal* 2002, **368**(Pt 2):397-404.
 121. Fry B: **Venomous reptiles and their toxins: evolution, pathophysiology and biodiscovery:** Oxford University Press; 2015.
 122. Richet G: **The discovery of anaphylaxis, a brief but triumphant encounter of two physiologists (1902).** *Histoire des sciences médicales* 2003, **37**(4):463-469.
 123. Richet C: **Anaphylaxis.** Translated by Murray Bligh. *Liverpool: University* 1913.
 124. Portier P, Richet C: **Sur les effets physiologiques du poison des filaments pêcheurs et des tentacules des Coelentérés (hypnotoxine).** *CR Acad Sci (Paris)* 1902, **134**:247.
 125. Russo AJ, Calton GJ, Burnett JW: **The relationship of the possible allergic response to jellyfish envenomation and serum antibody titers.** *Toxicon* 1983, **21**(4):475-480.
 126. Chi V, Pennington MW, Norton RS, Tarcha EJ, Londono LM, Sims-Fahey B, Upadhyay SK, Lakey JT, Iadonato S, Wulff H: **Development of a sea anemone toxin as an immunomodulator for therapy of autoimmune diseases.** *Toxicon* 2012, **59**(4):529-546.
 127. Suput D: **Interactions of cnidarian toxins with the immune system.** *Inflammation & Allergy-Drug Targets (Formerly Current Drug Targets-Inflammation & Allergy)* 2011, **10**(5):429-437.
 128. Ruëff F, Przybilla B, Biló MB, Müller U, Scheipl F, Aberer W, Birnbaum J, Bodzenta-Lukaszyk A, Bonifazi F, Bucher C: **Predictors of severe systemic anaphylactic reactions in patients with Hymenoptera venom allergy: importance of baseline serum tryptase—a study of the European Academy of Allergology and Clinical Immunology Interest Group on Insect Venom Hypersensitivity.** *Journal of Allergy and Clinical Immunology* 2009, **124**(5):1047-1054.
 129. Müller U, Johansen N, Petersen A, Fromberg-Nielsen J, Haeberli G: **Hymenoptera venom allergy: analysis of double positivity to honey bee and *Vespula* venom by estimation of IgE antibodies to species-specific major allergens Api m1 and Ves v5.** *Allergy* 2009, **64**(4):543-548.
 130. Bonadonna P, Zanotti R, Pagani M, Caruso B, Perbellini O, Colarossi S, Olivieri E, Dama A, Schiappoli M, Senna G: **How much specific is the association between hymenoptera venom allergy and mastocytosis?** *Allergy* 2009, **64**(9):1379-1382.
 131. Niedoszytko M, De Monchy J, Van Doormaal J, Jassem E, Oude Elberink J: **Mastocytosis and insect venom allergy: diagnosis, safety and efficacy of venom immunotherapy.** *Allergy* 2009, **64**(9):1237-1245.

132. Menez A: **Perspectives in molecular toxinology**: John Wiley & Sons; 2002.
133. Lewis RJ, Garcia ML: **Therapeutic potential of venom peptides**. *Nature Reviews Drug Discovery* 2003, **2**(10):790-802.
134. Ménez A, Gillet D, Grishin E: **Toxins: threats and benefits**. *Recent Research Developments in Toxins from Bacteria and Other Organisms* 2006:1-33.
135. Koh CY, Kini RM: **From snake venom toxins to therapeutics—cardiovascular examples**. *Toxicon* 2012, **59**(4):497-506.
136. Norton R, Pennington M, Beeton C: **Venoms to drugs: Venom as a source for the development of human therapeutics**. 2013.
137. King G: **Venoms to drugs: venom as a source for the development of human therapeutics**: Royal Society of Chemistry; 2015.
138. Hayashi MA, Camargo AC: **The Bradykinin-potentiating peptides from venom gland and brain of Bothrops jararaca contain highly site specific inhibitors of the somatic angiotensin-converting enzyme**. *Toxicon* 2005, **45**(8):1163-1170.
139. Bray GM: **Exenatide**. *American journal of health-system pharmacy* 2006, **63**(5).
140. Dardevet L, Rani D, Aziz TAE, Bazin I, Sabatier J-M, Fadl M, Brambilla E, De Waard M: **Chlorotoxin: a helpful natural scorpion peptide to diagnose glioma and fight tumor invasion**. *Toxins* 2015, **7**(4):1079-1101.
141. Mamelak AN, Jacoby DB: **Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601)**. *Expert opinion on drug delivery* 2007, **4**(2):175-186.
142. Castañeda O, Sotolongo V, Amor AM, Stöcklin R, Anderson AJ, Harvey AL, Engström Å, Wernstedt C, Karlsson E: **Characterization of a potassium channel toxin from the Caribbean Sea anemone Stichodactyla helianthus**. *Toxicon* 1995, **33**(5):603-613.
143. Norton R, Pennington M, Beeton C: **Transforming a toxin into a therapeutic: the sea anemone potassium channel blocker ShK toxin for treatment of autoimmune diseases**. In.: Royal Society of Chemistry: London, UK; 2015.
144. Leyendeckers H, Voth E, Schicha H, Hunzelmann N, Banga P, Schmitz J: **Frequent detection of thyroid peroxidase-specific IgG+ memory B cells in blood of patients with autoimmune thyroid disease**. *European journal of immunology* 2002, **32**(11):3126-3132.
145. Beeton C, Pennington MW, Wulff H, Singh S, Nugent D, Crossley G, Khaytin I, Calabresi PA, Chen C-Y, Gutman GA: **Targeting effector memory T cells with a selective peptide inhibitor of Kv1. 3 channels for therapy of autoimmune diseases**. *Molecular pharmacology* 2005, **67**(4):1369-1381.
146. Arcangeli A, Crociani O, Lastraioli E, Masi A, Pillozzi S, Becchetti A: **Targeting ion channels in cancer: a novel frontier in antineoplastic therapy**. *Current medicinal chemistry* 2009, **16**(1):66-93.
147. Cossins D: **From Toxins to Therapeutics**. In: *The Scientist*. <http://www.the-scientist.com/?articles.view/articleNo/34745/title/From-Toxins-to-Therapeutics/>; 2013.
148. Lavergne V, King GF, Lewis RJ, Alewood PF: **Peptide Therapeutics from Venomous Creatures**. In: *Pain Therapeutics*. 2013: 217-246.
149. Escoubas P, Sollod B, King GF: **Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach**. *Toxicon* 2006, **47**(6):650-663.
150. Wood DL, Miljenović T, Cai S, Raven RJ, Kaas Q, Escoubas P, Herzig V, Wilson D, King GF: **ArachnoServer: a database of protein toxins from spiders**. *BMC genomics* 2009, **10**(1):375.
151. Herzig V, Wood DL, Newell F, Chaumeil P-A, Kaas Q, Binford GJ, Nicholson GM, Gorse D, King GF: **ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures**. *Nucleic acids research* 2011, **39**(suppl_1):D653-D657.
152. Maček P, Lebez D: **Isolation and characterization of three lethal and hemolytic toxins from the sea anemone Actinia equina L**. *Toxicon* 1988, **26**(5):441-451.

153. Martin EJ: **Observations on the toxic sea anemone, *Rhodactis howesii* (Coelenterata).** 1960.
154. Bernheimer AW, Avigad LS: **Toxins of the sea anemone *Epiactis prolifera*.** *Archives of biochemistry and biophysics* 1982, **217**(1):174-180.
155. Frazão B, Antunes A: **Jellyfish bioactive compounds: methods for wet-lab work.** *Marine drugs* 2016, **14**(4):75.
156. Fautin DG: **Structural diversity, systematics, and evolution of cnidae.** *Toxicon* 2009, **54**(8):1054-1064.
157. Fautin DG, Malarky L, Soberon J: **Latitudinal diversity of sea anemones (Cnidaria: Actiniaria).** *The Biological Bulletin* 2013, **224**(2):89-98.
158. Rodríguez AA, Cassoli JS, Sa F, Dong ZQ, de Freitas JC, Pimenta AM, de Lima ME, Konno K, Lee SMY, Garateix A: **Peptide fingerprinting of the neurotoxic fractions isolated from the secretions of sea anemones *Stichodactyla helianthus* and *Bunodosoma granulifera*. New members of the APETx-like family identified by a 454 pyrosequencing approach.** *Peptides* 2012, **34**(1):26-38.
159. Kintner AH, Seymour JE, Edwards SL: **Variation in lethality and effects of two Australian chirodropid jellyfish venoms in fish.** *Toxicon* 2005, **46**(6):699-708.
160. Feng J, Yu H, Li C, Xing R, Liu S, Wang L, Cai S, Li P: **Isolation and characterization of lethal proteins in nematocyst venom of the jellyfish *Cyanea nozakii* Kishinouye.** *Toxicon* 2010, **55**(1):118-125.
161. Lee H, Jung E-s, Kang C, Yoon WD, Kim J-S, Kim E: **Scyphozoan jellyfish venom metalloproteinases and their role in the cytotoxicity.** *Toxicon* 2011, **58**(3):277-284.
162. Nagai H, Takuwa K, Nakao M, Sakamoto B, Crow GL, Nakajima T: **Isolation and characterization of a novel protein toxin from the Hawaiian box jellyfish (sea wasp) *Carybdea alata*.** *Biochemical and biophysical research communications* 2000, **275**(2):589-594.
163. Nagai H, Takuwa K, Nakao M, Ito E, Miyake M, Noda M, Nakajima T: **Novel proteinaceous toxins from the box jellyfish (sea wasp) *Carybdea rastoni*.** *Biochemical and biophysical research communications* 2000, **275**(2):582-588.
164. Nagai H, Takuwa-Kuroda K, Nakao M, OSHIRO N, IWANAGA S, NAKAJIMA T: **A novel protein toxin from the deadly box jellyfish (sea wasp, Habu-kurage) *Chiropsalmus quadrigatus*.** *Bioscience, biotechnology, and biochemistry* 2002, **66**(1):97-102.
165. Xiao L, Liu S, He Q, Wang Q, Ye X, Liu G, Nie F, Zhao J, Zhang L: **The acute toxicity and hematological characterization of the effects of tentacle-only extract from the jellyfish *Cyanea capillata*.** *Marine drugs* 2011, **9**(4):526-534.
166. Li R, Yu H, Xue W, Yue Y, Liu S, Xing R, Li P: **Jellyfish venomomics and venom gland transcriptomics analysis of *Stomolophus meleagris* to reveal the toxins associated with sting.** *Journal of Proteomics* 2014, **106**:17-29.
167. Li R, Yu H, Xing R, Liu S, Qing Y, Li K, Li B, Meng X, Cui J, Li P: **Application of nanoLC-MS/MS to the shotgun proteomic analysis of the nematocyst proteins from jellyfish *Stomolophus meleagris*.** *Journal of Chromatography B* 2012, **899**:86-95.
168. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical biochemistry* 1976, **72**(1-2):248-254.
169. Brinkman D, Burnell J: **Identification, cloning and sequencing of two major venom proteins from the box jellyfish, *Chironex fleckeri*.** *Toxicon* 2007, **50**(6):850-860.
170. Weston AJ, Chung R, Dunlap WC, Morandini AC, Marques AC, Moura-da-Silva AM, Ward M, Padilla G, da Silva LF, Andreakis N: **Proteomic characterisation of toxins isolated from nematocysts of the South Atlantic jellyfish *Olindias sambaquiensis*.** *Toxicon* 2013, **71**:11-17.

171. Ponce D, López-Vera E, Aguilar MB, Sánchez-Rodríguez J: **Preliminary results of the in vivo and in vitro characterization of a tentacle venom fraction from the jellyfish *Aurelia aurita*.** *Toxins* 2013, **5**(12):2420-2433.
172. Kim E, Lee S, Kim J-S, Yoon WD, Lim D, Hart AJ, Hodgson WC: **Cardiovascular effects of *Nemopilema nomurai* (Scyphozoa: Rhizostomeae) jellyfish venom in rats.** *Toxicology letters* 2006, **167**(3):205-211.
173. Badré S: **Bioactive toxins from stinging jellyfish.** *Toxicon* 2014, **91**:114-125.
174. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ: **Protein measurement with the Folin phenol reagent.** *J biol Chem* 1951, **193**(1):265-275.
175. Waddell WJ: **A simple ultraviolet spectrophotometric method for the determination of protein.** *Journal of Laboratory and Clinical Medicine* 1956, **48**:311-314.
176. Rodriguez AA, Ständker L, Zaharenko AJ, Garateix AG, Forssmann W-G, Béress L, Valdés O, Hernandez Y, Laguna A: **Combining multidimensional liquid chromatography and MALDI-TOF-MS for the fingerprint analysis of secreted peptides from the unexplored sea anemone species *Phymanthus crucifer*.** *Journal of Chromatography B* 2012, **903**:30-39.
177. Zaharenko AJ, Ferreira Jr WA, Oliveira JS, Richardson M, Pimenta DC, Konno K, Portaro FCV, de Freitas JC: **Proteomics of the neurotoxic fraction from the sea anemone *Bunodosoma cangicum* venom: Novel peptides belonging to new classes of toxins.** *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 2008, **3**(3):219-225.
178. Gusmani L, Avian M, Galil B, Patriarca P, Rottini G: **Biologically active polypeptides in the venom of the jellyfish *Rhopilema nomadica*.** *Toxicon* 1997, **35**(5):637-648.
179. Ayed Y, Bousabbah M, Mabrouk HB, Morjen M, Marrakchi N, Bacha H: **Impairment of the cell-to-matrix adhesion and cytotoxicity induced by the Mediterranean jellyfish *Pelagia noctiluca* venom and its fractions in cultured glioblastoma cells.** *Lipids in health and disease* 2012, **11**(1):84.
180. Sánchez-Rodríguez J, Lucio-Martínez N: **Isolation and prepurification of active compounds in venom from *Pelagia noctiluca* (Scyphozoa: Pelagiidae) from the Caribbean Sea Aislamiento y prepurificación de los compuestos activos presentes en el veneno de *Pelagia noctiluca* (Scyphozoa: Pelagiidae) del Mar Caribe.** *Ciencias Marinas* 2011, **37**(3):369-377.
181. Li R, Yu H, Feng J, Xing R, Liu S, Wang L, Qin Y, Li K, Li P: **Two-step purification and in vitro characterization of a hemolysin from the venom of jellyfish *Cyanea nozakii* Kishinouye.** *International journal of biological macromolecules* 2011, **49**(1):14-19.
182. Chaousis S, Smout M, Wilson D, Loukas A, Mulvenna J, Seymour J: **Rapid short term and gradual permanent cardiotoxic effects of vertebrate toxins from *Chironex fleckeri* (Australian box jellyfish) venom.** *Toxicon* 2014, **80**:17-26.
183. Brinkman D, Burnell J: **Partial purification of cytolytic venom proteins from the box jellyfish, *Chironex fleckeri*.** *Toxicon* 2008, **51**(5):853-863.
184. Winter KL, Isbister GK, McGowan S, Konstantakopoulos N, Seymour JE, Hodgson WC: **A pharmacological and biochemical examination of the geographical variation of *Chironex fleckeri* venom.** *Toxicology letters* 2010, **192**(3):419-424.
185. Gerber F, Krummen M, Potgieter H, Roth A, Siffrin C, Spoendlin C: **Practical aspects of fast reversed-phase high-performance liquid chromatography using 3µm particle packed columns and monolithic columns in pharmaceutical development and production working under current good manufacturing practice.** *Journal of Chromatography A* 2004, **1036**(2):127-133.
186. Brown P, Hartwick RA: **High performance liquid chromatography.** 1988.
187. Barnes J: **High performance liquid chromatography:** John Wiley & Sons; 1992.
188. Dorsey JG, Dill KA: **The molecular mechanism of retention in reversed-phase liquid chromatography.** *Chemical Reviews* 1989, **89**(2):331-346.

189. Howard G, Martin A: **The separation of the C12-C18 fatty acids by reversed-phase partition chromatography.** *Biochemical Journal* 1950, **46**(5):532.
190. Songping L: **Protocols for peptidomic analysis of spider venoms.** *Peptidomics: methods and protocols* 2010:75-85.
191. Yuan C, Jin Q, Tang X, Hu W, Cao R, Yang S, Xiong J, Xie C, Xie J, Liang S: **Proteomic and peptidomic characterization of the venom from the Chinese bird spider, *Ornithoctonus huwena* Wang.** *Journal of proteome research* 2007, **6**(7):2792-2801.
192. Newton KA, Clench MR, Deshmukh R, Jeyaseelan K, Strong PN: **Mass fingerprinting of toxic fractions from the venom of the Indian red scorpion, *Mesobuthus tamulus*: biotope-specific variation in the expression of venom peptides.** *Rapid communications in mass spectrometry* 2007, **21**(21):3467-3476.
193. Nascimento DG, Rates B, Santos DM, Verano-Braga T, Barbosa-Silva A, Dutra AA, Biondi I, Martin-Eauclaire MF, De Lima ME, Pimenta AM: **Moving pieces in a taxonomic puzzle: venom 2D-LC/MS and data clustering analyses to infer phylogenetic relationships in some scorpions from the Buthidae family (*Scorpiones*).** *Toxicon* 2006, **47**(6):628-639.
194. Anderson NL, Anderson NG: **Proteome and proteomics: new technologies, new concepts, and new words.** *Electrophoresis* 1998, **19**(11):1853-1861.
195. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I: **Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*.** *Electrophoresis* 1995, **16**(1):1090-1094.
196. Swinbanks D: **Australia backs innovation, shuns telescope.** 1995.
197. Downard KM: **Francis William Aston: the man behind the mass spectrograph.** *European Journal of Mass Spectrometry* 2007, **13**(3):177-190.
198. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM: **Electrospray ionization—principles and practice.** *Mass Spectrometry Reviews* 1990, **9**(1):37-70.
199. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T: **Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.** *Rapid communications in mass spectrometry* 1988, **2**(8):151-153.
200. Cottrell JS, London U: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *electrophoresis* 1999, **20**(18):3551-3567.
201. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification.** *Nature biotechnology* 2008, **26**(12):1367-1372.
202. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M: **A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.** *Nature protocols* 2009, **4**(5):698-705.
203. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G: **PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.** *Rapid communications in mass spectrometry* 2003, **17**(20):2337-2342.
204. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I: **UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view.** *Plant Bioinformatics: Methods and Protocols* 2016:23-54.
205. Jungo F, Bougueleret L, Xenarios I, Poux S: **The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data.** *Toxicon* 2012, **60**(4):551-557.
206. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S: **Database resources of the national center for biotechnology information.** *Nucleic acids research* 2011, **39**(suppl 1):D38-D51.

207. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *ELECTROPHORESIS* 1999, **20**(18):3551-3567.
208. Seymour SL, Hunter C: **ProteinPilot™ Software Overview.** In: *High quality, in-depth protein identification and protein expression analysis.* 2015.
209. Rossi DT, Sinz M: **Mass spectrometry in drug discovery:** CRC Press; 2001.
210. Kang C, Han D-y, Park K-I, Pyo M-J, Heo Y, Lee H, Kim GS, Kim E: **Characterization and neutralization of Nemopilema nomurai (Scyphozoa: Rhizostomeae) jellyfish venom using polyclonal antibody.** *Toxicon* 2014, **86**:116-125.
211. Calvete JJ, Juárez P, Sanz L: **Snake venomomics. Strategy and applications.** *Journal of Mass Spectrometry* 2007, **42**(11):1405-1414.
212. Escoubas P, Quinton L, Nicholson GM: **Venomomics: unravelling the complexity of animal venoms with mass spectrometry.** *Journal of mass spectrometry* 2008, **43**(3):279-295.
213. Cottrell JS: **Protein identification using MS/MS data.** *Journal of proteomics* 2011, **74**(10):1842-1851.
214. Burlingame AL, Carr SA: **Mass spectrometry in the biological sciences:** Springer Science & Business Media; 2012.
215. Gold V, Loening K, McNaught A, Shemi P: **IUPAC compendium of chemical terminology.** *Blackwell Science, Oxford* 1997.
216. Li R, Yu H, Xing R, Liu S, Qing Y, Li K, Li B, Meng X, Cui J, Li P: **Isolation and in vitro partial characterization of hemolytic proteins from the nematocyst venom of the jellyfish Stomolophus meleagris.** *Toxicology in Vitro* 2013, **27**(6):1620-1625.
217. Moran Y, Praher D, Schlesinger A, Ayalon A, Tal Y, Technau U: **Analysis of soluble protein contents from the nematocysts of a model sea anemone sheds light on venom evolution.** *Marine Biotechnology* 2013, **15**(3):329-339.
218. Cassoli JS, Verano-Braga T, Oliveira JS, Montandon GG, Cologna CT, Peigneur S, de Castro Pimenta AM, Kjeldsen F, Roepstorff P, Tytgat J: **The proteomic profile of Stichodactyla duerdeni secretion reveals the presence of a novel O-linked glycopeptide.** *Journal of proteomics* 2013, **87**:89-102.
219. Goudet C, Ferrer T, Galàn L, Artiles A, Batista CF, Possani LD, Alvarez J, Aneiros A, Tytgat J: **Characterization of two Bunodosoma granulifera toxins active on cardiac sodium channels.** *British journal of pharmacology* 2001, **134**(6):1195-1206.
220. Alves P, Arnold RJ, Novotny MV, Radivojac P, Reilly JP, Tang H: **Advancement in protein inference from shotgun proteomics using peptide detectability.** In: *Pacific Symposium on Biocomputing: 2007;* 2007: 409-420.
221. Wolters DA, Washburn MP, Yates JR: **An automated multidimensional protein identification technology for shotgun proteomics.** *Analytical chemistry* 2001, **73**(23):5683-5690.
222. Hu L, Ye M, Jiang X, Feng S, Zou H: **Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis—a review.** *Analytica chimica acta* 2007, **598**(2):193-204.
223. Coon JJ: **Collisions or electrons? Protein sequence analysis in the 21st century.** In: ACS Publications; 2009.
224. Brinkman DL, Aziz A, Loukas A, Potriquet J, Seymour J, Mulvenna J: **Venom proteome of the box jellyfish Chironex fleckeri.** *PloS one* 2012, **7**(12):e47866.
225. Brinkman DL, Jia X, Potriquet J, Kumar D, Dash D, Kvaskoff D, Mulvenna J: **Transcriptome and venom proteome of the box jellyfish Chironex fleckeri.** *BMC genomics* 2015, **16**(1):407.
226. Baggerman G, Vierstraete E, De Loof A, Schoofs L: **Gel-based versus gel-free proteomics: a review.** *Combinatorial chemistry & high throughput screening* 2005, **8**(8):669-677.

227. Craig R, Beavis RC: **TANDEM: matching proteins with tandem mass spectra.** *Bioinformatics* 2004, **20**(9):1466-1467.
228. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
229. Campos A, Apraiz I, da Fonseca RR, Cristobal S: **Shotgun analysis of the marine mussel *Mytilus edulis* hemolymph proteome and mapping the innate immunity elements.** *Proteomics* 2015, **15**(23-24):4021-4029.
230. Campos A, Danielsson G, Farinha AP, Kuruvilla J, Warholm P, Cristobal S: **Shotgun proteomics to unravel marine mussel (*Mytilus edulis*) response to long-term exposure to low salinity and propranolol in a Baltic Sea microcosm.** *Journal of proteomics* 2016, **137**:97-106.
231. de Magalhães JP, Finch CE, Janssens G: **Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions.** *Ageing research reviews* 2010, **9**(3):315-323.
232. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.** *BMC genomics* 2012, **13**(1):341.
233. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M: **Comparison of next-generation sequencing systems.** *BioMed Research International* 2012, **2012**.
234. Ducancel F, Durban J, Verdenaud M: **Transcriptomics and venomics: implications for medicinal chemistry.** *Future medicinal chemistry* 2014, **6**(15):1629-1643.
235. Huang Y-F, Chen S-C, Chiang Y-S, Chen T-H, Chiu K-P: **Palindromic sequence impedes sequencing-by-ligation mechanism.** *BMC systems biology* 2012, **6**(2):S10.
236. Metzker ML: **Emerging technologies in DNA sequencing.** *Genome research* 2005, **15**(12):1767-1776.
237. Coupland P, Chandra T, Quail M, Reik W, Swerdlow H: **Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation.** *Biotechniques* 2012, **53**(6).
238. Biosciences P: **Pacific Biosciences introduces new chemistry with longer read lengths to detect novel features in DNA sequence and advance genome studies of large organisms.** Website <http://investor.pacificbiosciences.com/releasedetail.cfm> 2013.
239. Rhoads A, Au KF: **PacBio sequencing and its applications.** *Genomics, proteomics & bioinformatics* 2015, **13**(5):278-289.
240. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ: **The methylomes of six bacteria.** *Nucleic acids research* 2012:gks891.
241. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE: **Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.** *Nature methods* 2013, **10**(6):563-569.
242. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin C-S, Iliopoulos D: **Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany.** *New England Journal of Medicine* 2011, **365**(8):709-717.
243. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F: **Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing.** *BMC genomics* 2011, **12**(1):245.
244. Maxmen A: **Transcriptomics for the Animal Kingdom.** In: *The Scientist*. <http://www.the-scientist.com/?articles.view/articleNo/36132/title/Transcriptomics-for-the-Animal-Kingdom/>; 2013.

245. Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Välimäki N, Paulin L, Kvist J, Wahlberg N: **The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera.** *Nature communications* 2014, **5**.
246. Wheat CW, Fescemyer HW, Kvist J, Tas E, Vera JC, Frilander MJ, Hanski I, Marden JH: **Functional genomics of life history variation in a butterfly metapopulation.** *Molecular Ecology* 2011, **20**(9):1813-1828.
247. Durban J, Pérez A, Sanz L, Gómez A, Bonilla F, Chacón D, Sasa M, Angulo Y, Gutiérrez JM, Calvete JJ: **Integrated “omics” profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*.** *BMC genomics* 2013, **14**(1):234.
248. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nature biotechnology* 2011, **29**(7):644-652.
249. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels.** *Bioinformatics* 2012, **28**(8):1086-1092.
250. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome research* 1999, **9**(9):868-877.
251. Corrêa-Netto C, de LM Junqueira-de-Azevedo I, Silva DA, Ho PL, Leitão-de-Araújo M, Alves MLM, Sanz L, Foguel D, Zingali RB, Calvete JJ: **Snake venomomics and venom gland transcriptomic analysis of Brazilian coral snakes, *Micrurus altirostris* and *M. corallinus*.** *Journal of proteomics* 2011, **74**(9):1795-1809.
252. Durban J, Juárez P, Angulo Y, Lomonte B, Flores-Díaz M, Alape-Girón A, Sasa M, Sanz L, Gutiérrez JM, Dopazo J: **Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing.** *BMC genomics* 2011, **12**(1):259.
253. Fry BG, Scheib H, de Azevedo IdLJ, Silva DA, Casewell NR: **Novel transcripts in the maxillary venom glands of advanced snakes.** *Toxicon* 2012, **59**(7):696-708.
254. Hill RE, Mackessy SP: **Characterization of venom (Duvernoy’s secretion) from twelve species of colubrid snakes and partial sequence of four venom proteins.** *Toxicon* 2000, **38**(12):1663-1687.
255. McGivern JJ, Wray KP, Margres MJ, Couch ME, Mackessy SP, Rokyta DR: **RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes.** *BMC genomics* 2014, **15**(1):1061.
256. Junqueira-de-Azevedo IL, Campos PF, Ching AT, Mackessy SP: **Colubrid venom composition: an-omics perspective.** *Toxins* 2016, **8**(8):230.
257. Rokyta DR, Lemmon AR, Margres MJ, Aronow K: **The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*).** *BMC genomics* 2012, **13**(1):312.
258. Rozman KK, Doull J: **Dose and time as variables of toxicity.** *Toxicology* 2000, **144**(1):169-178.
259. Rozman KK, Doull J: **General principles of toxicology.** *Environmental Toxicology* 1998:1-11.
260. Repetto M: **Toxicología Fundamental: Toxicología Experimental. Toxicidad Aguda.** In.: ENPSES-MERCIE GROUP ^ eEspaña España; 2002.
261. MacKENZIE RD, Byerrum RU, Decker CF, Hoppert CA, LANGHAM RF, Mich E: **Chronic toxicity studies.** In: *II Hexavalent and trivalent chromium administered in drinking water to rats.* vol. 18; 1958: 232-234.
262. CYTED PIdC, para el Desarrollo T: **Manual de Técnicas de Investigación.** Editado por CYTED 1995.
263. Oecd: **OECD Guidelines for the Testing of Chemicals:** Organization for Economic; 1994.
264. Edström H, Gyllensporre D, Hallams E, Ratti L, Zyla B, Heller F, Gillingham J, Matlár JH, Petersson M, Sloan SR: **Organisation for Economic Co-operation and Development (OECD).** 2016.

265. Woodward R: **The organisation for economic Co-operation and development (OECD)**: Routledge; 2009.
266. Russell WMS, Burch RL: **The principles of humane experimental technique**. *The principles of humane experimental technique* 1959.
267. Lilienblum W, Dekant W, Foth H, Gebel T, Hengstler J, Kahl R, Kramer P-J, Schweinfurth H, Wollin K-M: **Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH)**. *Archives of toxicology* 2008, **82**(4):211-236.
268. Höfer T, Gerner I, Gundert-Remy U, Liebsch M, Schulte A, Spielmann H, Vogel R, Wettig K: **Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation**. *Archives of toxicology* 2004, **78**(10):549-564.
269. Festing S, Wilkinson R: **The ethics of animal research. Talking Point on the use of animals in scientific research**. *EMBO Reports* 2007, **8**(6):526-530.
270. Irwin S: **Prediction of drug effects from animals to man**. In: *Ciba Foundation Symposium-Animal Behaviour and Drug Action: 1964*: Wiley Online Library; 1964: 269-285.
271. Irwin S: **Drug screening and evaluation of new compounds in animals**. *Animal and clinical pharmacologic techniques in drug evaluation* 1964:36-54.
272. Schaumann W, Stoepel K: **Zur quantitativen Beurteilung von zentraler Erregung und Dämpfung im Tierversuch**. *Naunyn-Schmiedeberg's Archives of Pharmacology* 1961, **241**(3):383-392.
273. Geyer MA: **Approaches to the characterization of drug effects on locomotor activity in rodents**. *Testing and Evaluation of Drugs of Abuse* New York: Wiley-Liss 1990:81-99.
274. Ramos A, Berton O, Mormède P, Chaoulloff F: **A multiple-test study of anxiety-related behaviours in six inbred rat strains**. *Behavioural brain research* 1997, **85**(1):57-69.
275. Bourin M, Hascoët M: **The mouse light/dark box test**. *European journal of pharmacology* 2003, **463**(1):55-65.
276. Boguszewski P, Zagrodzka J: **Emotional changes related to age in rats—a behavioral analysis**. *Behavioural brain research* 2002, **133**(2):323-332.
277. Hall C, Ballachey EL: **A study of the rat's behavior in a field. A contribution to method in comparative psychology**. *University of California Publications in Psychology* 1932.
278. Crunelli V, Bernasconi S: **A new device to measure different size movements: Studies on d-amphetamine-induced locomotion and stereotypy**. *Journal of Pharmacological Methods* 1979, **2**(1):43-50.
279. Kanji N, Owen JA: **A microprocessor-based animal monitoring system**. *Journal of pharmacological methods* 1980, **3**(1):71-82.
280. Walsh RN, Cummins RA: **The open-field test: A critical review**. *Psychological bulletin* 1976, **83**(3):482.
281. Carola V, D'Olimpio F, Brunamonti E, Mangia F, Renzi P: **Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice**. *Behavioural brain research* 2002, **134**(1):49-57.
282. File SE: **What can be learned from the effects of benzodiazepines on exploratory behavior?** *Neuroscience & Biobehavioral Reviews* 1985, **9**(1):45-54.
283. Boyce-Rustay JM, Holmes A: **Genetic inactivation of the NMDA receptor NR2A subunit has anxiolytic-and antidepressant-like effects in mice**. *Neuropsychopharmacology* 2006, **31**(11):2405-2414.
284. Deacon R, Koros E, Bornemann K, Rawlins J: **Aged Tg2576 mice are impaired on social memory and open field habituation tests**. *Behavioural brain research* 2009, **197**(2):466-468.

285. Tsujimura A, Matsuki M, Takao K, Yamanishi K, Miyakawa T, Hashimoto-Gotoh T: **Mice lacking the kf-1 gene exhibit increased anxiety-but not despair-like behavior.** *Frontiers in behavioral neuroscience* 2008, **2**:4.
286. Kim DH, Hung TM, Bae KH, Jung JW, Lee S, Yoon BH, Cheong JH, Ko KH, Ryu JH: **Gomisin A improves scopolamine-induced memory impairment in mice.** *European Journal of Pharmacology* 2006, **542**(1):129-135.
287. Yang MH, Yoon KD, Chin Y-W, Park JH, Kim SH, Kim YC, Kim J: **Neuroprotective effects of *Dioscorea opposita* on scopolamine-induced memory impairment in in vivo behavioral tests and in vitro assays.** *Journal of ethnopharmacology* 2009, **121**(1):130-134.
288. Kim DH, Yoon BH, Kim Y-W, Lee S, Shin BY, Jung JW, Kim HJ, Lee YS, Choi JS, Kim SY: **The seed extract of *Cassia obtusifolia* ameliorates learning and memory impairments induced by scopolamine or transient cerebral hypoperfusion in mice.** *Journal of pharmacological sciences* 2007, **105**(1):82-93.
289. Morris R: **Developments of a water-maze procedure for studying spatial learning in the rat.** *Journal of neuroscience methods* 1984, **11**(1):47-60.
290. Vorhees CV, Williams MT: **Morris water maze: procedures for assessing spatial and related forms of learning and memory.** *Nature protocols* 2006, **1**(2):848-858.
291. Li X, Blizzard KK, Zeng Z, DeVries AC, Hurn PD, McCullough LD: **Chronic behavioral testing after focal ischemia in the mouse: functional recovery and the effects of gender.** *Experimental neurology* 2004, **187**(1):94-104.
292. WoRMS: **Cnidaria.** In. World Register of Marine Species at <http://www.marinespecies.org/aphia.php?p=taxdetails&id=1267> (accessed on 24 May 2017).
293. Frazão B, Campos A, Osório H, Thomas B, Leandro S, Teixeira A, Vasconcelos V, Antunes A: **Analysis of *Pelagia noctiluca* proteome Reveals a Red Fluorescent Protein, a Zinc Metalloproteinase and a Peroxiredoxin.** *The Protein Journal* 2017, **36**(2):77-97.
294. Frazão B: **A Genomic and Proteomic Study of Sea Anemones and Jellyfish from Portugal.** 2017.
295. Turk T, Kem WR: **The phylum Cnidaria and investigations of its toxins and venoms until 1990.** *Toxicon* 2009, **54**(8):1031-1037.
296. Iturralde-Vinent MA: **Ensayo sobre la paleogeografía del Cuaternario de Cuba.** In: *V Congreso de Geología y Minería: 2003*; 2003: 54-74.
297. Ching AT, Rocha MM, Paes Leme AF, Pimenta DC, de Fátima D Furtado M, Serrano SM, Ho PL, Junqueira-de-Azevedo IL: **Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome.** *FEBS letters* 2006, **580**(18):4417-4422.
298. Gonçalves-Machado L, Pla D, Sanz L, Jorge RJB, Leitão-De-Araújo M, Alves MLM, Alvares DJ, De Miranda J, Nowatzki J, de Moraes-Zani K: **Combined venomomics, venom gland transcriptomics, bioactivities, and antivenomics of two *Bothrops jararaca* populations from geographic isolated regions within the Brazilian Atlantic rainforest.** *Journal of proteomics* 2016, **135**:73-89.
299. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes.** *Toxicon* 2011, **57**(5):657-671.
300. Henderson RW, Powell R: **Natural history of West Indian reptiles and amphibians:** University Press of Florida; 2009.
301. Uetz P, Hošek J: **The Reptile Database.** In. <http://www.reptile-database.org>, accessed on June 10, 2017; 2013.
302. Kardong KV: **The evolution of the venom apparatus in snakes from colubrids to viperids and elapids.** *Mem Inst Butantan* 1982, **46**:105-118.
303. Kardong KV: **Colubrid snakes and Duvernoy's "venom" glands.** *Journal of Toxicology: Toxin Reviews* 2002, **21**(1-2):1-19.

304. Fry BG, Wüster W, Ramjan R, Fadil S, Jackson T, Martelli P, Kini RM: **Analysis of Colubroidea snake venoms by liquid chromatography with mass spectrometry: evolutionary and toxinological implications.** *Rapid Communications in Mass Spectrometry* 2003, **17**(18):2047-2062.
305. Mackessy SP, Saviola AJ: **Understanding biological roles of venoms among the Caenophidia: The importance of rear-fanged snakes.** *Integrative and Comparative Biology* 2016:icw110.
306. Fry BG, Lumsden NG, Wüster W, Wickramaratna JC, Hodgson WC, Kini RM: **Isolation of a neurotoxin (α -colubritoxin) from a nonvenomous colubrid: evidence for early origin of venom in snakes.** *Journal of molecular evolution* 2003, **57**(4):446-452.
307. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SR, Kuruppu S, Fung K, Hedges SB, Richardson MK: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439**(7076):584-588.
308. Fry BG, Wroe S, Teeuwisse W, van Osch MJ, Moreno K, Ingle J, McHenry C, Ferrara T, Clausen P, Scheib H: **A central role for venom in predation by *Varanus komodoensis* (Komodo Dragon) and the extinct giant *Varanus* (*Megalania*) *priscus*.** *Proceedings of the National Academy of Sciences* 2009, **106**(22):8969-8974.
309. Saint Girons H: **Les glandes céphaliques exocrines des reptiles. I: Données anatomiques et histologiques.** In: *Annales des sciences naturelles Zoologie et biologie animale: 1988*: Elsevier; 1988: 221-255.
310. Halpern M: **Nasal chemical senses in reptiles: structure and function.** Pp 423–523 in Gans C, Crews D (eds) *Biology of the Reptilia, Vol. 18, Brain. Hormones, and Behavior* Chicago/IL: University of Chicago Press Google Scholar 1992.
311. Payne A: **The harderian gland: a tercentennial review.** *Journal of anatomy* 1994, **185**(Pt 1):1.
312. Chieffi G, Baccari GC, Di Matteo L, d'Istria M, Minucci S, Varriale B: **Cell biology of the Harderian gland.** *International review of cytology* 1996, **168**:1-80.
313. Daly M, Brugler MR, Cartwright P, Collins AG, Dawson MN, Fautin DG, France SC, McFadden CS, Opresko DM, Rodriguez E: **The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus.** 2007.
314. Watson GM, Mariscal RN: **Ultrastructure and sulfur cytochemistry of nematocyst development in catch tentacles of the sea anemone *Haliplanella luciae* (Cnidaria: Anthozoa).** *Journal of ultrastructure research* 1984, **87**(2):159-171.
315. Weill R: **Contribution à l'étude des cnidaires et de leurs nématocystes. I. Recherches sur les nématocystes (morphologie, physiologie, développement)...[II. Valeur taxonomique du cnidome.]:** Laboratoire d'évolution des êtres organisés; 1934.
316. Nagai H: **Marine protein toxins.** In: *Handbook of marine natural products.* Springer; 2012: 1388-1419.
317. Oliveira JS, Fuentes-Silva D, King GF: **Development of a rational nomenclature for naming peptide and protein toxins from sea anemones.** *Toxicon* 2012, **60**(4):539-550.
318. Nesher N, Shapira E, Sher D, Moran Y, Tsveyer L, Turchetti-Maia AL, Horowitz M, Hochner B, Zlotkin E: **AdE-1, a new inotropic Na⁺ channel toxin from *Aiptasia diaphana*, is similar to, yet distinct from, known anemone Na⁺ channel toxins.** *Biochemical Journal* 2013, **451**(1):81-90.
319. Orts DJ, Peigneur S, Madio B, Cassoli JS, Montandon GG, Pimenta A, Bicudo JE, Freitas JC, Zaharenko AJ, Tytgat J: **Biochemical and electrophysiological characterization of two sea anemone type 1 potassium toxins from a geographically distant population of *Bunodosoma caissarum*.** *Marine drugs* 2013, **11**(3):655-679.
320. Gleibs S, Mebs D, Werdning B: **Studies on the origin and distribution of palytoxin in a Caribbean coral reef.** *Toxicon* 1995, **33**(11):1531-1537.

321. Villar RM, Gil-Longo J, Daranas AH, Souto MaL, Fernandez JJ, Peixinho S, Barral MA, Santafe G, Rodríguez J, Jiménez C: **Evaluation of the effects of several zoanthamine-type alkaloids on the aggregation of human platelets.** *Bioorganic & medicinal chemistry* 2003, **11**(10):2301-2306.
322. Lakshmi V, Saxena A, Pandey K, Bajpai P, Misra-Bhattacharya S: **Antifilarial activity of Zoanthus species (Phylum Coelenterata, Class Anthzoa) against human lymphatic filaria, Brugia malayi.** *Parasitology research* 2004, **93**(4):268-273.
323. Santos Y, Martínez M, Sandoval A, Rodríguez AA, Falcón A, de la Coteria EPH, Aguilar MB, Flores P, Felix R, Arreguín R: **Arrhythmogenic effect of a crude extract from sea anemone Condylactis gigantea: Possible involvement of rErg1 channels.** *Toxicon* 2013, **67**:47-54.
324. Leal-García M, García-Ortuño L, Ruiz-Azuara L, Gracia-Mora I, Luna-delVillar J, Sumano H: **Assessment of acute respiratory and cardiovascular toxicity of casiopeinas in anaesthetized dogs.** *Basic & clinical pharmacology & toxicology* 2007, **101**(3):151-158.
325. Benitah J-P, Alvarez JL, Gómez AM: **L-type Ca²⁺ current in ventricular cardiomyocytes.** *Journal of molecular and cellular cardiology* 2010, **48**(1):26-36.
326. Chambers DJ: **Mechanisms and alternative methods of achieving cardiac arrest.** *The Annals of thoracic surgery* 2003, **75**(2):S661-S666.
327. Newton CR, Delgado JH, Gomez HF: **Calcium and beta receptor antagonist overdose: a review and update of pharmacological principles and management.** In: *Seminars in respiratory and critical care medicine: 2002*: Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662; 2002: 019-026.
328. Ojetti V, Migneco A, Bononi F, De Lorenzo A, Silveri NG: **Calcium channel blockers, beta-blockers and digitalis poisoning: Management in the emergency room.** *European review for medical and pharmacological sciences* 2005, **9**(4):241.
329. Levine M, Boyer EW, Pozner CN, Geib A-J, Thomsen T, Mick N, Thomas SH: **Assessment of hyperglycemia after calcium channel blocker overdoses involving diltiazem or verapamil.** *Critical care medicine* 2007, **35**(9):2071-2075.
330. Yan L, Herrington J, Goldberg E, Dulski PM, Bugianesi RM, Slaughter RS, Banerjee P, Brochu RM, Priest BT, Kaczorowski GJ: **Stichodactyla helianthus peptide, a pharmacological tool for studying Kv3. 2 channels.** *Molecular pharmacology* 2005, **67**(5):1513-1521.
331. Tejuca M, Diaz I, Figueredo R, Roque L, Pazos F, Martinez D, Iznaga-Escobar N, Perez R, Alvarez C, Lanio M: **Construction of an immunotoxin with the pore forming protein Stl and ior C5, a monoclonal antibody against a colon cancer cell line.** *International immunopharmacology* 2004, **4**(6):731-744.
332. Wu CH: **Palytoxin: membrane mechanisms of action.** *Toxicon* 2009, **54**(8):1183-1189.
333. Deeds JR, Handy SM, White KD, Reimer JD: **Palytoxin found in Palythoa sp. zoanthids (Anthozoa, Hexacorallia) sold in the home aquarium trade.** *PLoS One* 2011, **6**(4):e18235.
334. Lazcano-Pérez F, Vivas O, Román-González SA, Rodríguez-Bustamante E, Castro H, Arenas I, García DE, Sánchez-Puig N, Arreguín-Espinosa R: **A purified Palythoa venom fraction delays sodium current inactivation in sympathetic neurons.** *Toxicon* 2014, **82**:112-116.
335. Huang C, MORLIGHEM J-ER, Hefeng Z, Erica Patrícia L, GOMES PB, PEREZ CD, Inchio L, LEE SMY, Radis-Baptista G: **Basal repertoire of toxin-related transcripts in the zoanthid Protopalycha variabilis (cnidaria, anthozoa).** *CJPT* 2015, **29**(S1):104-105.
336. Kelecom A: **Studies of Brazilian Marine Invertebrates. VIII. Zoanthosterol, a New Sterol from the Zoanthid Zoanthus Sociatus (Hexacorallia, Zoanthidea).** *Bulletin des Sociétés Chimiques Belges* 1981, **90**(9):971-976.

337. Fukuzawa S, Hayashi Y, Uemura D, Nagatsu A, Yamada K, Ijuin Y: **The isolation and structures of five new alkaloids, norzoanthamine, oxyzoanthamine, norzoanthaminone, cyclozoanthamine and epinorzoanthamine.** *Heterocycl Commun* 1995, **1**:207-214.
338. Cen-Pacheco F, Martín MN, Fernández JJ, Hernández Daranas A: **New oxidized zoanthamines from a Canary Islands *Zoanthus* sp.** *Marine drugs* 2014, **12**(10):5188-5196.
339. Domínguez-Pérez D, Díaz-García CM, García-Delgado N, Sierra-Gómez Y, Castañeda O, Antunes A: **Insights into the toxicological properties of a low molecular weight fraction from *Zoanthus sociatus* (Cnidaria).** *Marine drugs* 2013, **11**(8):2873-2881.
340. Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, Kuchеров G: **NORINE: a database of nonribosomal peptides.** *Nucleic acids research* 2008, **36**(suppl 1):D326-D331.
341. Caboche S, Leclere V, Pupin M, Kuchеров G, Jacques P: **Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity.** *Journal of bacteriology* 2010, **192**(19):5143-5150.
342. Miller MA, Kudela RM, Mekebre A, Crane D, Oates SC, Tinker MT, Staedler M, Miller WA, Toy-Choutka S, Dominik C: **Evidence for a novel marine harmful algal bloom: cyanotoxin (microcystin) transfer from land to sea otters.** *PLoS One* 2010, **5**(9):e12576.
343. Chen DZ, Boland MP, Smillie MA, Klix H, Ptak C, Andersen RJ, Holmes CF: **Identification of protein phosphatase inhibitors of the microcystin class in the marine environment.** *Toxicon* 1993, **31**(11):1407-1414.
344. Lehman P, Boyer G, Hall C, Waller S, Gehrts K: **Distribution and toxicity of a new colonial *Microcystis aeruginosa* bloom in the San Francisco Bay Estuary, California.** *Hydrobiologia* 2005, **541**(1):87-99.
345. Matthiensen A, Beattie KA, Yunes JS, Kaya K, Codd GA: **[d-Leu 1] Microcystin-LR, from the cyanobacterium *Microcystis* RST 9501 and from a *Microcystis* bloom in the Patos Lagoon estuary, Brazil.** *Phytochemistry* 2000, **55**(5):383-387.
346. Tonk L, Bosch K, Visser PM, Huisman J: **Salt tolerance of the harmful cyanobacterium *Microcystis aeruginosa*.** *Aquatic Microbial Ecology* 2007, **46**(2):117-123.
347. Richardson LL, Sekar R, Myers JL, Gantar M, Voss JD, Kaczmarzsky L, Remily ER, Boyer GL, Zimba PV: **The presence of the cyanobacterial toxin microcystin in black band disease of corals.** *FEMS microbiology letters* 2007, **272**(2):182-187.
348. Ferranti P, Fabbrocino S, Nasi A, Caira S, Bruno M, Serpe L, Gallo P: **Liquid chromatography coupled to quadruple time-of-flight tandem mass spectrometry for microcystin analysis in freshwaters: method performances and characterisation of a novel variant of microcystin-RR.** *Rapid Communications in Mass Spectrometry* 2009, **23**(9):1328-1336.
349. Mbukwa EA, Msagati TA, Mamba BB: **Quantitative variations of intracellular microcystin-LR,RR and-YR in samples collected from four locations in Hartbeespoort Dam in North West Province (South Africa) during the 2010/2011 summer season.** *International journal of environmental research and public health* 2012, **9**(10):3484-3505.
350. Namikoshi M, Sun F, Choi BW, Rinehart KL, Carmichael WW, Evans WR, Beasley VR: **Seven More Microcystins from Homer Lake Cells: Application of the General Method for Structure Assignment of Peptides Containing. alpha., beta.-Dehydroamino Acid Unit (s).** *The Journal of Organic Chemistry* 1995, **60**(12):3671-3679.
351. Vasas G, Szydlowska D, Gáspár A, Welker M, Trojanowicz M, Borbély G: **Determination of microcystins in environmental samples using capillary electrophoresis.** *Journal of biochemical and biophysical methods* 2006, **66**(1):87-97.

352. Bogialli S, Bruno M, Curini R, Di Corcia A, Laganá A, Mari B: **Simple Assay for Analyzing Five Microcystins and Nodularin in Fish Muscle Tissue: Hot Water Extraction Followed by Liquid Chromatography– Tandem Mass Spectrometry.** *Journal of agricultural and food chemistry* 2005, **53**(17):6586-6592.
353. Namikoshi M, Yuan M, Sivonen K, Carmichael WW, Rinehart KL, Rouhiainen L, Sun F, Brittain S, Otsuki A: **Seven new microcystins possessing two L-glutamic acid units, isolated from *Anabaena* sp. strain 186.** *Chemical research in toxicology* 1998, **11**(2):143-149.
354. Stoner RD, Adams WH, Slatkin DN, Siegelman HW: **The effects of single L-amino acid substitutions on the lethal potencies of the microcystins.** *Toxicon* 1989, **27**(7):825-828.
355. Sivonen K, Namikoshi M, Evans W, Carmichael W, Sun F, Rouhiainen L, Luukkainen R, Rinehart K: **Isolation and characterization of a variety of microcystins from seven strains of the cyanobacterial genus *Anabaena*.** *Applied and environmental microbiology* 1992, **58**(8):2495-2500.
356. Meriluoto J, Sandström A, Eriksson J, Remaud G, Graig AG, Chattopadhyaya J: **Structure and toxicity of a peptide hepatotoxin from the cyanobacterium *Oscillatoria agardhii*.** *Toxicon* 1989, **27**(9):1021-1034.
357. Namikoshi M, Rinehart KL, Sakai R, Stotts RR, Dahlem AM, Beasley VR, Carmichael WW, Evans WR: **Identification of 12 hepatotoxins from a Homer Lake bloom of the cyanobacteria *Microcystis aeruginosa*, *Microcystis viridis*, and *Microcystis wesenbergii*: nine new microcystins.** *The Journal of organic chemistry* 1992, **57**(3):866-872.
358. Sivonen K, Namikoshi M, Evans WR, Fardig M, Carmichael WW, Rinehart KL: **Three new microcystins, cyclic heptapeptide hepatotoxins, from *Nostoc* sp. strain 152.** *Chemical research in toxicology* 1992, **5**(4):464-469.
359. Namikoshi M, Sivonen K, Evans WR, Carmichael WW, Rouhiainen L, Luukkainen R, Rinehart KL: **Structures of three new homotyrosine-containing microcystins and a new homophenylalanine variant from *Anabaena* sp. strain 66.** *Chemical research in toxicology* 1992, **5**(5):661-666.
360. Liyanage MR, Bakshi K, Volkin DB, Middaugh CR: **Ultraviolet Absorption Spectroscopy of Peptides.** *Therapeutic Peptides: Methods and Protocols* 2014:225-236.
361. Scopes R: **Measurement of protein by spectrophotometry at 205 nm.** *Analytical biochemistry* 1974, **59**(1):277-282.
362. Anthis NJ, Clore GM: **Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm.** *Protein Science* 2013, **22**(6):851-858.
363. Gill SC, Von Hippel PH: **Calculation of protein extinction coefficients from amino acid sequence data.** *Analytical biochemistry* 1989, **182**(2):319-326.
364. Zhang T, Kiyonami R, Wang L, Jiang G: **Identification and Quantitation of Microcystins by Targeted Full-Scan LC-MS/MS.** In: *Application Note 569*. Edited by Thermo-Fisher-Scientific. San Jose, CA, USA; 2016.
365. Neumann U, Campos V, Cantarero S, Urrutia H, Heinze R, Weckesser J, Erhard M: **Co-occurrence of non-toxic (cyauropeptolin) and toxic (microcystin) peptides in a bloom of *Microcystis* sp. from a Chilean lake.** *Systematic and applied microbiology* 2000, **23**(2):191-197.
366. Tsuji K, Watanuki T, Kondo F, Watanabe MF, Suzuki S, Nakazawa H, Suzuki M, Uchida H, Harada K-I: **Stability of microcystins from cyanobacteria—II. Effect of UV light on decomposition and isomerization.** *Toxicon* 1995, **33**(12):1619-1631.
367. Martin C, Oberer L, Ino T, König W, Busch M, Weckesser J: **Cyauropeptolins, new depsipeptides from the cyanobacterium *Microcystis* sp. PCC 7806.** *The Journal of antibiotics* 1993, **46**(10):1550.
368. Carneiro RL, Dörr FA, Dörr F, Bortoli S, Delherbe N, Vásquez M, Pinto E: **Co-occurrence of microcystin and microginin congeners in Brazilian strains of *Microcystis* sp.** *FEMS microbiology ecology* 2012, **82**(3):692-702.

369. Puddick J, Prinsep MR, Wood SA, Kaufononga SA, Cary SC, Hamilton DP: **High levels of structural diversity observed in microcystins from Microcystis CAWBG11 and characterization of six new microcystin congeners.** *Marine drugs* 2014, **12**(11):5372-5395.
370. Taton A, Grubisic S, Balthasart P, Hodgson DA, Laybourn-Parry J, Wilmotte A: **Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes.** *FEMS microbiology ecology* 2006, **57**(2):272-289.
371. Fumanti B, Cavacini P, Alfinito S: **Benthic algal mats of some lakes of Inexpressible Island (northern Victoria Land, Antarctica).** *Polar Biology* 1996, **17**(1):25-30.
372. Timo N: **Microcystin congeners described in the literature.** Available online: https://figshare.com/articles/Microcystin_congeners_described_in_the_literature/880756 (accessed on 26/01/2017).
373. Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA: **Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system.** *Chemistry & biology* 2000, **7**(10):753-764.
374. Carmichael W: **Cyanobacteria secondary metabolites—the cyanotoxins.** *Journal of applied bacteriology* 1992, **72**(6):445-459.
375. Zohary T: **Hyperscums of the cyanobacterium Microcystis aeruginosa in a hypertrophic lake (Hartbeespoort Dam, South Africa).** *Journal of Plankton Research* 1985, **7**(3):399-409.
376. Ueno Y, Nagata S, Tsutsumi T, Hasegawa A, Yoshida F, Suttajit M, Mebs D, Pütsch M, Vasconcelos V: **Survey of microcystins in environmental water by a highly sensitive immunoassay based on monoclonal antibody.** *Natural toxins* 1996, **4**(6):271-276.
377. Lopes VR, Ramos V, Martins A, Sousa M, Welker M, Antunes A, Vasconcelos VM: **Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries.** *Marine environmental research* 2012, **73**:7-16.
378. Frazão B, Martins R, Vasconcelos V: **Are known cyanotoxins involved in the toxicity of picoplanktonic and filamentous North Atlantic marine cyanobacteria?** *Marine drugs* 2010, **8**(6):1908-1919.
379. Landsberg JH: **The effects of harmful algal blooms on aquatic organisms.** *Reviews in Fisheries Science* 2002, **10**(2):113-390.
380. DeMott WR, Moxter F: **Foraging cyanobacteria by copepods: responses to chemical defense and resource abundance.** *Ecology* 1991, **72**(5):1820-1834.
381. Malbrouck C, Kestemont P: **Effects of microcystins on fish.** *Environmental Toxicology and Chemistry* 2006, **25**(1):72-86.
382. Amorim Á, Vasconcelos Vt: **Dynamics of microcystins in the mussel Mytilus galloprovincialis.** *Toxicon* 1999, **37**(7):1041-1052.
383. Williams DE, Dawe SC, Kent ML, Andersen RJ, Craig M, Holmes CF: **Bioaccumulation and clearance of microcystins from salt water mussels, Mytilus edulis, and in vivo evidence for covalently bound microcystins in mussel tissues.** *Toxicon* 1997, **35**(11):1617-1625.
384. Vasconcelos V, Oliveira S, Teles FO: **Impact of a toxic and a non-toxic strain of Microcystis aeruginosa on the crayfish Procambarus clarkii.** *Toxicon* 2001, **39**(10):1461-1470.
385. Zimba PV, Camus A, Allen EH, Burkholder JM: **Co-occurrence of white shrimp, Litopenaeus vannamei, mortalities and microcystin toxin in a southeastern USA shrimp facility.** *Aquaculture* 2006, **261**(3):1048-1055.
386. Backer LC, Carmichael W, Kirkpatrick B, Williams C, Irvin M, Zhou Y, Johnson TB, Nierenberg K, Hill VR, Kieszak SM: **Recreational exposure to low concentrations of microcystins during an algal bloom in a small lake.** *Marine Drugs* 2008, **6**(2):389-406.

387. De Silva ED, Williams DE, Andersen RJ, Klix H, Holmes CF, Allen TM: **Motuporin, a potent protein phosphatase inhibitor isolated from the Papua New Guinea sponge *Theonella swinhoei* Gray.** *Tetrahedron letters* 1992, **33**(12):1561-1564.
388. Rinehart KL, Namikoshi M, Choi BW: **Structure and biosynthesis of toxins from blue-green algae (cyanobacteria).** *Journal of applied phycology* 1994, **6**(2):159-176.
389. Krishnamurthy T, Carmichael W, Sarver E: **Toxic peptides from freshwater cyanobacteria (blue-green algae). I. Isolation, purification and characterization of peptides from *Microcystis aeruginosa* and *Anabaena flos-aquae*.** *Toxicon* 1986, **24**(9):865-873.
390. Zhao Y, Shi K, Su X, Xie L, Yan Y: **Microcystin-LR induces dysfunction of insulin secretion in rat insulinoma (INS-1) cells: Implications for diabetes mellitus.** *Journal of hazardous materials* 2016, **314**:11-21.
391. Strohm M, Hassman M, Kořata B, Kolděek M: **mMass data miner: an open source alternative for mass spectrometric data analysis.** *Rapid Communications in Mass Spectrometry* 2008, **22**(6):905-908.
392. Strohm M, Kavan D, Novak P, Volny M, Havlicek V: **mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data.** *Analytical chemistry* 2010, **82**(11):4648-4651.
393. Orts DJ, Moran Y, Cologna CT, Peigneur S, Madio B, Praher D, Quinton L, Pauw E, Bicudo JE, Tytgat J: **BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker.** *FEBS Journal* 2013, **280**(19):4839-4852.
394. Urbarova I, Karlsen BO, Okkenhaug S, Seternes OM, Johansen SD, Emblem Å: **Digital Marine Bioprospecting: Mining New Neurotoxin Drug Candidates from the Transcriptomes of Cold-Water Sea Anemones.** *Marine Drugs* 2012, **10**(10):2265-2279.
395. Macrander J, Brugler MR, Daly M: **A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps.** *BMC Genomics* 2015, **16**(1):221.
396. Norton RS: **Structure and structure-function relationships of sea anemone proteins that interact with the sodium channel.** *Toxicon* 1991, **29**(9):1051-1084.
397. Fautin DG, (2015): **Hexacorallians of the World. *Bunodactis verrucosa*.** In. Fautin, Daphne G. (2013). Hexacorallians of the World. Accessed through: World Register of Marine Species at <http://www.marinespecies.org/aphia.php?p=taxdetails&id=100819> on 2017-05-12.
398. Campos A, Puerto M, Prieto A, Cameán A, Almeida AM, Coelho AV, Vasconcelos V: **Protein extraction and two-dimensional gel electrophoresis of proteins in the marine mussel *Mytilus galloprovincialis*: an important tool for protein expression studies, food quality and safety assessment.** *Journal of the Science of Food and Agriculture* 2013, **93**(7):1779-1787.
399. Neuhoﬀ V, Arold N, Taube D, Ehrhardt W: **Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250.** *Electrophoresis* 1988, **9**(6):255-262.
400. Osorio H, Reis CA: **Mass spectrometry methods for studying glycosylation in cancer.** *Mass Spectrometry Data Analysis in Proteomics* 2013:301-316.
401. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M: **UniProt: the universal protein knowledgebase.** *Nucleic acids research* 2004, **32**(suppl_1):D115-D119.
402. Consortium U: **UniProt: the universal protein knowledgebase.** *Nucleic acids research* 2017, **45**(D1):D158-D169.
403. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis.** *Nature methods* 2009, **6**(5):359.
404. Jungo F, Estreicher A, Bairoch A, Bougueleret L, Xenarios I: **Animal toxins: how is complexity represented in databases?** *Toxins* 2010, **2**(2):262-282.

405. Jungo F, Bairoch A: **Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase.** *Toxicon* 2005, **45**(3):293-301.
406. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *science* 2007, **317**(5834):86-94.
407. Zhuang Y, Hou H, Zhao X, Zhang Z, Li B: **Effects of collagen and collagen hydrolysate from jellyfish (*Rhopilema esculentum*) on mice skin photoaging induced by UV irradiation.** *Journal of Food Science* 2009, **74**(6):H183-H188.
408. Puerto M, Campos A, Prieto A, Cameán A, de Almeida AM, Coelho AV, Vasconcelos V: **Differential protein expression in two bivalve species; *Mytilus galloprovincialis* and *Corbicula fluminea*; exposed to *Cylindrospermopsis raciborskii* cells.** *Aquatic toxicology* 2011, **101**(1):109-116.
409. Zhang X: **Less is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-level Proteomics.** *Molecular & Cellular Proteomics* 2015, **14**(9):2441-2453.
410. Wu CC, MacCoss MJ, Howell KE, Yates JR: **A method for the comprehensive proteomic analysis of membrane proteins.** *Nature biotechnology* 2003, **21**(5):532-538.
411. Ji H, Wang J, Guo J, Li Y, Lian S, Guo W, Yang H, Kong F, Zhen L, Guo L: **Progress in the biological function of alpha-enolase.** *Animal Nutrition* 2016, **2**(1):12-17.
412. Iida H, Yahara I: **Yeast heat-shock protein of Mr 48,000 is an isoprotein of enolase.** 1985.
413. Butterworth P: **Book review: Ribonucleases: structure and function. G. D'Alessio and JF Riordan (Eds). Academic Press, New York and London. 670+ xix pages (1997).** In.: Wiley Online Library; 1998.
414. Davenport RC, Bash PA, Seaton BA, Karplus M, Petsko GA, Ringe D: **Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: an analog of the intermediate on the reaction pathway.** *Biochemistry* 1991, **30**(24):5821-5826.
415. Cooper SJ, Leonard GA, McSweeney SM, Thompson AW, Naismith JH, Qamar S, Plater A, Berry A, Hunter WN: **The crystal structure of a class II fructose-1, 6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold.** *Structure* 1996, **4**(11):1303-1315.
416. Horecker B, Tsolas O, Lai C: **6 Aldolases.** *The enzymes* 1972, **7**:213-258.
417. Harrison PM, Arosio P: **The ferritins: molecular properties, iron storage function and cellular regulation.** *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1996, **1275**(3):161-203.
418. Arosio P, Levi S: **Ferritin, iron homeostasis, and oxidative damage 1, 2.** *Free Radical Biology and Medicine* 2002, **33**(4):457-463.
419. Kim K, Kim I, Lee K-Y, Rhee S, Stadtman E: **The isolation and purification of a specific" protector" protein which inhibits enzyme inactivation by a thiol/Fe (III)/O₂ mixed-function oxidation system.** *Journal of Biological Chemistry* 1988, **263**(10):4704-4711.
420. Rhee SG, Woo HA, Kil IS, Bae SH: **Peroxiredoxin functions as a peroxidase and a regulator and sensor of local peroxides.** *Journal of Biological Chemistry* 2012, **287**(7):4403-4410.
421. Cui H, Wang Y, Wang Y, Qin S: **Genome-wide analysis of putative peroxiredoxin in unicellular and filamentous cyanobacteria.** *BMC evolutionary biology* 2012, **12**(1):220.
422. Maksimenko AV, Vavaev AV: **Antioxidant enzymes as potential targets in cardioprotection and treatment of cardiovascular diseases. Enzyme antioxidants: the next stage of pharmacological counterwork to the oxidative stress.** *Heart international* 2012, **7**(1):14-19.

423. Ruan Z, Liu G, Wang B, Zhou Y, Lu J, Wang Q, Zhao J, Zhang L: **First report of a peroxiredoxin homologue in jellyfish: molecular cloning, expression and functional characterization of CcPrx4 from Cyanea capillata.** *Marine drugs* 2014, **12**(1):214-231.
424. Fukai T, Ushio-Fukai M: **Superoxide dismutases: role in redox signaling, vascular function, and diseases.** *Antioxidants & redox signaling* 2011, **15**(6):1583-1606.
425. Culma MF, Pereanez JA, Rangel VN, Lomonte B: **Snake venomomics of Bothrops punctatus, a semiarboreal pitviper species from Antioquia, Colombia.** *PeerJ* 2014, **2**:e246.
426. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates III JR: **Protein analysis by shotgun/bottom-up proteomics.** *Chemical reviews* 2013, **113**(4):2343-2394.
427. Keller BO, Sui J, Young AB, Whittall RM: **Interferences and contaminants encountered in modern mass spectrometry.** *Analytica chimica acta* 2008, **627**(1):71-81.
428. Daly M, Gusmão LC, Reft AJ, Rodríguez E: **Phylogenetic signal in mitochondrial and nuclear markers in sea anemones (Cnidaria, Actiniaria).** *Integrative and Comparative Biology* 2010, **50**(3):371-388.
429. Moffatt BA, Ashihara H: **Purine and pyrimidine nucleotide synthesis and metabolism.** *The Arabidopsis Book* 2002:e0018.
430. Lonsdale D: **A review of the biochemistry, metabolism and clinical benefits of thiamin (e) and its derivatives.** *Evidence-based Complementary and Alternative Medicine* 2006, **3**(1):49-59.
431. Borbón H, Váldez S, Alvarado-Mesén J, Soto R, Vega I: **Antimicrobial properties of sea anemone Anthopleura nigrescens from Pacific coast of Costa Rica.** *Asian Pacific Journal of Tropical Biomedicine* 2016, **6**(5):418-421.
432. Rocha J, Peixe L, Gomes N, Calado R: **Cnidarians as a source of new marine bioactive compounds—An overview of the last decade and future steps for bioprospecting.** *Marine drugs* 2011, **9**(10):1860-1886.
433. Mariottini GL, Grice ID: **Antimicrobials from Cnidarians. A New Perspective for Anti-Infective Therapy?** *Marine drugs* 2016, **14**(3):48.
434. Davy SK, Allemand D, Weis VM: **Cell biology of cnidarian-dinoflagellate symbiosis.** *Microbiology and Molecular Biology Reviews* 2012, **76**(2):229-261.
435. Bellis ES, Howe DK, Denver DR: **Genome-wide polymorphism and signatures of selection in the symbiotic sea anemone Aiptasia.** *BMC genomics* 2016, **17**(1):160.
436. Tsai IH, Wang YM, Chiang TY, Chen YL, Huang RJ: **Purification, cloning and sequence analyses for pro-metalloprotease-disintegrin variants from Deinagkistrodon acutus venom and subclassification of the small venom metalloproteases.** *European Journal of Biochemistry* 2000, **267**(5):1359-1367.
437. Undheim EA, Sunagar K, Herzig V, Kely L, Low DH, Jackson TN, Jones A, Kurniawan N, King GF, Ali SA: **A proteomics and transcriptomics investigation of the venom from the barychelid spider Trittame loki (brush-foot trapdoor).** *Toxins* 2013, **5**(12):2488-2503.
438. Sarras MP, Li Y, Leontovich A, Zhang JS: **Structure, expression, and developmental function of early divergent forms of metalloproteinases in hydra.** *Cell research* 2002, **12**(3):163-176.
439. Leontovich AA, Zhang J, Shimokawa K-i, Nagase H, Sarras M: **A novel hydra matrix metalloproteinase (HMMP) functions in extracellular matrix degradation, morphogenesis and the maintenance of differentiated cells in the foot process.** *Development* 2000, **127**(4):907-920.
440. Yan L, Fei K, Zhang J, Dexter S, Sarras M: **Identification and characterization of hydra metalloproteinase 2 (HMP2): a meprin-like astacin metalloproteinase that functions in foot morphogenesis.** *Development* 2000, **127**(1):129-141.

441. Pan T-L, Gröger H, Schmid V, Spring J: **A toxin homology domain in an astacin-like metalloproteinase of the jellyfish *Podocoryne carnea* with a dual role in digestion and development.** *Development genes and evolution* 1998, **208**(5):259-266.
442. Tsai I-H, Wang Y-M, Cheng AC, Starkov V, Osipov A, Nikitin I, Makarova Y, Ziganshin R, Utkin Y: **cDNA cloning, structural, and functional analyses of venom phospholipases A 2 and a Kunitz-type protease inhibitor from steppe viper *Vipera ursinii renardi*.** *Toxicon* 2011, **57**(2):332-341.
443. Valentin E, Ghomashchi F, Gelb MH, Lazdunski M, Lambeau G: **Novel human secreted phospholipase A2 with homology to the group III bee venom enzyme.** *Journal of Biological Chemistry* 2000, **275**(11):7492-7496.
444. Fry BG, Wüster W: **Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences.** *Molecular biology and evolution* 2004, **21**(5):870-883.
445. Komori Y, Nikai T, Sugihara H: **Purification and characterization of a lethal toxin from the venom of *Heloderma horridum*.** *Biochemical and biophysical research communications* 1988, **154**(2):613-619.
446. Fry BG, Roelants K, Winter K, Hodgson WC, Griesman L, Kwok HF, Scanlon D, Karas J, Shaw C, Wong L: **Novel venom proteins produced by differential domain-expression strategies in beaded lizards and gila monsters (genus *Heloderma*).** *Molecular Biology and Evolution* 2009:msp251.
447. Jouiaei M, Yanagihara AA, Madio B, Nevalainen TJ, Alewood PF, Fry BG: **Ancient venom systems: a review on cnidaria toxins.** *Toxins* 2015, **7**(6):2251-2271.
448. Skewes M: ***Aulactinia verrucosa* Gem anemone.** In: Tyler-Walters H. and Hiscock K. (eds) Marine Life Information Network: Biology and Sensitivity Key Information Reviews, [on-line]. Plymouth: Marine Biological Association of the United Kingdom. Available from: <http://www.marlin.ac.uk/species/detail/1601>.
449. White B, (2004): ***Anthopleura xanthogrammica*.** In: Walla Walla University. at http://www.wallawalla.edu/academics/departments/biology/rosario/inverts/Cnidaria/CI-ass-Anthozoa/Subclass Zoantharia/Order Actiniaria/Anthopleura_xanthogrammica.html on 2017-05-12.
450. Sebens KP: **The allometry of feeding, energetics, and body size in three sea anemone species.** *The Biological Bulletin* 1981, **161**(1):152-171.
451. Hand C, Uhlinger KR: **The culture, sexual and asexual reproduction, and growth of the sea anemone *Nematostella vectensis*.** *The Biological Bulletin* 1992, **182**(2):169-176.
452. Cope WG, Bringolf RB, Buchwalter DB, Newton TJ, Ingersoll CG, Wang N, Augspurger T, Dwyer FJ, Barnhart MC, Neves RJ: **Differential exposure, duration, and sensitivity of unionoidean bivalve life stages to environmental contaminants.** *Journal of the North American Benthological Society* 2008, **27**(2):451-462.
453. Gilroy ÈA, Gillis PL, King LE, Bendo NA, Salerno J, Giacomini M, de Solla SR: **The effects of pharmaceuticals on a unionid mussel (*Lampsilis siliquoidea*): An examination of acute and chronic endpoints of toxicity across life stages.** *Environmental toxicology and chemistry* 2017.
454. Bleam DE, Couch KJ, Distler DA: **Key to the unionid mussels of Kansas.** *Transactions of the Kansas Academy of Science (1903)* 1999:83-91.
455. Ghirelli F: **Cephalotoxin: the crab-paralysing agent of the posterior salivary glands of cephalopods.** *Nature* 1959, **183**(4669):1192-1193.
456. McDonald N, Cottrell G: **Purification and mode of action of toxin from *Eledone cirrosa*.** *Comparative and general pharmacology* 1972, **3**(10):243-248.
457. Songdahl J, Shapiro B: **Purification and composition of a toxin from the posterior salivary gland of *Octopus dofleini*.** *Toxicon* 1974, **12**(2):109-112.

458. Cariello L, Zanetti L: **α -and β -cephalotoxin: two paralysing proteins from posterior salivary glands of *Octopus vulgaris*.** *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology* 1977, **57**(2):169-173.
459. McDowell SB: **Toxicocalamus, a New Guinea genus of snakes of the family Elapidae.** *Journal of Zoology* 1969, **159**(4):443-511.
460. Saint-Girons H: **Les glandes céphaliques exocrines des reptiles. II: Considérations fonctionnelles et évolutives.** In: *Annales des sciences naturelles Zoologie et biologie animale*: 1989: Elsevier; 1989: 1-17.
461. Dillemeijer P: **A comparative functional-anatomical study of the heads of some Viperidae;** 1959.
462. Harder J: **Glandula nova lachrymalis una cum ductu excretorio in cervis ed damis.** *Acta Erudit Publ Lipsias*:49-52.
463. Cowan FBM: **The ultrastructure of the lachrymal'salt'gland and the Harderian gland in the euryhaline *Malaclemys* and some closely related stenohaline emydines.** *Canadian journal of zoology* 1971, **49**(5):691-697.
464. Saint Girons H: **Histologie comparée des glandes orbitaires des Lépidosauriens.** *Ann Sci Nat Zool(Paris)* 1982, **4**:171-191.
465. Wetterberg L, Geller E, Yuwiler A: **Harderian gland: an extraretinal photoreceptor influencing the pineal gland in neonatal rats?** *Science* 1970, **167**(3919):884-885.
466. Reiter RJ, Richardson BA, Matthews SA, Lane SJ, Ferguson BN: **Rhythms in immunoreactive melatonin in the retina and Harderian gland of rats: persistence after pinealectomy.** *Life sciences* 1983, **32**(11):1229-1236.
467. Pevet P, Heth G, Hiam A, Nevo E: **Photoperiod perception in the blind mole rat (*Spalax ehrenbergi*, Nehring): involvement of the Harderian gland, atrophied eyes, and melatonin.** *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* 1984, **232**(1):41-50.
468. Young JA, Van Lennep EW: **The morphology of salivary glands:** Academic Press; 1978.
469. Walls GL: **The vertebrate eye and its adaptive radiation.** Oxford, England: Cranbrook Institute of Science; 1942.
470. Bellairs AA, Boyd J: **The Lachrymal Apparatus in Lizards and Snakes.—I. The Brille, the Orbital Glands, Lachrymal Canaliculi and Origin of the Lachrymal Duct.** In: *Proceedings of the Zoological Society of London*: 1947: Wiley Online Library; 1947: 81-108.
471. Rehorek SJ: **Squamate Harderian gland: an overview.** *The Anatomical Record* 1997, **248**(3):301-306.
472. Bentz E, Mason RT: **The role of the Harderian gland in the chemical ecology of garter snakes.** In: Department of Integrative Biology, Oregon State University (OSU); <http://masonlab.science.oregonstate.edu/>; 2017.
473. Savitsky AH: **The relationships of the xenodontine colubrid snakes related to *Ninia*;** 1974.
474. West G: **On the Histology of the Salivary, Buccal, and Harderian Glands of the Colubridae, with Notes on their Tooth-succession and the Relationships of the Poison-duct.** *Journal of the Linnean Society of London, Zoology* 1898, **26**(171):517-526.
475. Smith M, Bellairs AdA: **The Head Glands of Snakes, with remarks on the Evolution of the Parotid Gland and Teeth of the Opisthoglypha.** *Journal of the Linnean Society of London, Zoology* 1947, **41**(279):351-368.
476. Greene HW, Fogden M, Fogden P: **Snakes: the evolution of mystery in nature:** Univ of California Press; 2000.
477. Schwarz-Karsten H: **Über die Orbitaldrüsen von *Lacerta agilis*, *Lacerta muralis*, *Ophiosops elegans*, *Tarentola mauretanica* und *Tropidonotud natrix*.** *Morphol Jahrb* 1937, **80**:248-279.

478. Savitzky AH: **The origin of the New World proteroglyphous snakes and its bearing on the study of venom delivery systems in snakes.** University of Kansas; 1979.
479. McCarthy C: **Monophyly of elapid snakes (Serpentes: Elapidae). An assessment of the evidence.** *Zoological Journal of the Linnean Society* 1985, **83**(1):79-93.
480. Zaher H, de Oliveira L, Grazziotin FG, Campagner M, Jared C, Antoniazzi MM, Prudente AL: **Consuming viscous prey: a novel protein-secreting delivery system in neotropical snail-eating snakes.** *BMC evolutionary biology* 2014, **14**(1):58.
481. Pratt C: **The morphology of the ethmoidal region of Sphenodon and lizards.** In: *Proceedings of the zoological society of London: 1948*: Wiley Online Library; 1948: 171-201.
482. Parsons TS: **The nose and Jacobsons Organ.** In: *Biology of the Reptilia*. vol. 2. Morphology B. C. Gans, ed. : Academic Press, London; 1970.
483. McCleary RJ, Kini RM: **Non-enzymatic proteins from snake venoms: a gold mine of pharmacological tools and drug leads.** *Toxicon* 2013, **62**:56-74.
484. Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ, Chisholm SW: **Unlocking short read sequencing for metagenomics.** *PloS one* 2010, **5**(7):e11840.
485. council NR: **Guide for the care and use of laboratory animals:** National Academies Press; 2010.
486. Bregge-Silva C, Nonato MC, de Albuquerque S, Ho PL, de Azevedo ILJ, Diniz MRV, Lomonte B, Rucavado A, Díaz C, Gutiérrez JM: **Isolation and biochemical, functional and structural characterization of a novel L-amino acid oxidase from Lachesis muta snake venom.** *Toxicon* 2012, **60**(7):1263-1276.
487. Smit A, Hubley R, Green P: **RepeatMasker Open-4.0. 2013–2015.** *Institute for Systems Biology* <http://repeatmasker.org> 2015.
488. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
489. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nature methods* 2008, **5**(7):621-628.
490. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences.** *Bioinformatics* 2007, **23**(9):1073-1079.
491. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome biology* 2009, **10**(3):R25.
492. Chang Z, Wang Z, Li G: **The impacts of read length and transcriptome complexity for de novo assembly: A simulation study.** *PloS one* 2014, **9**(4):e94825.
493. Flower DR: **The lipocalin protein family: structure and function.** *Biochemical Journal* 1996, **318**(1):1-14.
494. Grzyb J, Latowski D, Strzałka K: **Lipocalins—a family portrait.** *Journal of plant physiology* 2006, **163**(9):895-915.
495. Miyawaki A, Matsushita F, Ryo Y, Mikoshiba K: **Possible pheromone-carrier function of two lipocalin proteins in the vomeronasal organ.** *The EMBO journal* 1994, **13**(24):5835.
496. Lee K-H, Wells RG, Reed RR: **Isolation of an olfactory cDNA: similarity to retinol-binding protein suggests a role in olfaction.** *Science* 1987, **235**:1053-1057.
497. Weldon CL, Mackessy SP: **Alsophinase, a new P-III metalloproteinase with α -fibrinogenolytic and hemorrhagic activity from the venom of the rear-fanged Puerto Rican Racer Alsophis portoricensis (Serpentes: Dipsadidae).** *Biochimie* 2012, **94**(5):1189-1198.
498. Terrat Y, Sunagar K, Fry BG, Jackson TN, Scheib H, Fourmy R, Verdenaud M, Blanchet G, Antunes A, Ducancel F: **Atractaspis aterrima toxins: The first insight**

- into the molecular evolution of venom in side-stabbers.** *Toxins* 2013, **5**(11):1948-1964.
499. Baumann K, Casewell NR, Ali SA, Jackson TN, Vetter I, Dobson JS, Cutmore SC, Nouwens A, Lavergne V, Fry BG: **A ray of venom: Combined proteomic and transcriptomic investigation of fish venom composition using barb tissue from the blue-spotted stingray (*Neotrygon kuhlii*).** *Journal of proteomics* 2014, **109**:188-198.
 500. Gewurz H, Zhang X-H, Lint TF: **Structure and function of the pentraxins.** *Current opinion in immunology* 1995, **7**(1):54-64.
 501. Andersen GR, Nissen P, Nyborg J: **Elongation factors in protein biosynthesis.** *Trends in biochemical sciences* 2003, **28**(8):434-441.
 502. Bum-Erdene K, Leffler H, Nilsson UJ, Blanchard H: **Structural characterisation of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3'-sulfo-lactose, and 2'-fucosyllactose.** *Scientific reports* 2016, **6**.
 503. Delacour D, Gouyer V, Zanetta J-P, Drobecq H, Leteurtre E, Grard G, Moreau-Hannedouche O, Maes E, Pons A, André S: **Galectin-4 and sulfatides in apical membrane trafficking in enterocyte-like cells.** *J Cell Biol* 2005, **169**(3):491-501.
 504. Stechly L, Morelle W, Dessein AF, André S, Grard G, Trinel D, Dejonghe MJ, Leteurtre E, Drobecq H, Trugnan G: **Galectin-4-regulated delivery of glycoproteins to the brush border membrane of enterocyte-like cells.** *Traffic* 2009, **10**(4):438-450.
 505. Muñiz M, Riezman H: **Trafficking of glycosylphosphatidylinositol anchored proteins from the endoplasmic reticulum to the cell surface.** *Journal of lipid research* 2016, **57**(3):352-360.
 506. Hernandez JD, Baum LG: **Ah, sweet mystery of death! Galectins and control of cell fate.** *Glycobiology* 2002, **12**(10):127R-136R.
 507. Morita T: **C-type lectin-related proteins from snake venoms.** *Current Drug Targets-Cardiovascular & Hematological Disorders* 2004, **4**(4):357-373.
 508. Morita T: **Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities.** *Toxicon* 2005, **45**(8):1099-1114.
 509. Andrault J-B, Gaillard I, Giorgi D, Rouquier S: **Expansion of the BPI family by duplication on human chromosome 20: characterization of the RY gene cluster in 20q11. 21 encoding olfactory transporters/antimicrobial-like peptides☆.** *Genomics* 2003, **82**(2):172-184.
 510. Olender T, Keydar I, Pinto JM, Tatarsky P, Alkelai A, Chien M-S, Fishilevich S, Restrepo D, Matsunami H, Gilad Y: **The human olfactory transcriptome.** *BMC genomics* 2016, **17**(1):619.
 511. Parker MR, Mason RT: **How to make a sexy snake: estrogen activation of female sex pheromone in male red-sided garter snakes.** *Journal of Experimental Biology* 2012, **215**(5):723-730.
 512. Brykczynska U, Tzika AC, Rodriguez I, Milinkovitch MC: **Contrasted evolution of the vomeronasal receptor repertoires in mammals and squamate reptiles.** *Genome biology and evolution* 2013, **5**(2):389-401.
 513. Dear T, Boehm T, Keverne E, Rabbitts T: **Novel genes for potential ligand-binding proteins in subregions of the olfactory mucosa.** *The EMBO journal* 1991, **10**(10):2813.
 514. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
 515. Consortium GO: **Gene ontology consortium: going forward.** *Nucleic acids research* 2015, **43**(D1):D1049-D1056.

516. Neill WT: **Evidence of Venom in Snakes of the Genera Alsophis and Rhadinaea.** *Copeia* 1954, **1954**(1):59-60.
517. Jaume M, Garrido O: **Notas sobre mordidas de jubo Alsophis (Serpentes: Colubridae).** *Miscelánea Zoológica* 1980, **11**:2-3.
518. Rodríguez-Cabrera TM, Torres J, Romero RM, Podio-Martínez J: **Predation Attempt by the Cuban Racer, *Cubophis cantherigerus* (Squamata: Dipsadidae) on the Cuban Giant Anole, *Anolis equestris buidei* (Squamata: Dactyloidae), a Threatened Endemic Subspecies.** *IRCF Reptiles and Amphibians* 2016, **23**:46–50.
519. Utkin YN: **Three-finger toxins, a deadly weapon of elapid venom—milestones of discovery.** *Toxicon* 2013, **62**:50-55.
520. Pawlak J, Mackessy SP, Fry BG, Bhatia M, Mourier G, Fruchart-Gaillard C, Servent D, Ménez R, Stura E, Ménez A: **Denmotoxin, a three-finger toxin from the colubrid snake *Boiga dendrophila* (Mangrove Catsnake) with bird-specific activity.** *Journal of Biological Chemistry* 2006, **281**(39):29030-29041.
521. Pawlak J, Mackessy SP, Sixberry NM, Stura EA, Le Du MH, Ménez R, Foo CS, Ménez A, Nirthanan S, Kini RM: **Irditoxin, a novel covalently linked heterodimeric three-finger toxin with high taxon-specific neurotoxicity.** *The FASEB Journal* 2009, **23**(2):534-545.
522. Sunagar K, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A: **Evolution of CRISPs associated with toxicoferan-reptilian venom and mammalian reproduction.** *Molecular biology and evolution* 2012, **29**(7):1807-1822.
523. Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJ, Van Osch MJ, Teeuwisse WM, Van Der Weerd L, Mcnaughtan JE, Kwok HF: **Functional and structural diversification of the Anguimorpha lizard venom system.** *Molecular & Cellular Proteomics* 2010, **9**(11):2369-2390.
524. Fry BG, Undheim EA, Ali SA, Jackson TN, Debono J, Scheib H, Ruder T, Morgenstern D, Cadwallader L, Whitehead D: **Squeezers and leaf-cutters: differential diversification and degeneration of the venom system in toxicoferan reptiles.** *Molecular & Cellular Proteomics* 2013, **12**(7):1881-1899.
525. Hargreaves AD, Swain MT, Logan DW, Mulley JF: **Testing the Toxicofera: comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system.** *Toxicon* 2014, **92**:140-156.
526. Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF: **Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins.** *Genome biology and evolution* 2014, **6**(8):2088-2095.
527. de Oliveira Junior NG, Franco OL: **Snake venoms: attractive antimicrobial proteinaceous compounds for therapeutic purposes.** *Cellular and molecular life sciences* 2013, **70**(24):4645-4658.
528. Samy RP, Gopalakrishnakone P, Chow VT, Ho B: **Viper metalloproteinase (*Agkistrodon halys pallas*) with antimicrobial activity against multi-drug resistant human pathogens.** *Journal of cellular physiology* 2008, **216**(1):54-68.
529. Castañeda O, Harvey AL: **Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels.** *Toxicon* 2009, **54**(8):1119-1124.
530. Niedermeyer TH, Strohm M: **mMass as a software tool for the annotation of cyclic peptide tandem mass spectra.** *PloS one* 2012, **7**(9):e44913.
531. De Hoffmann E, Stroobant V: **Mass spectrometry: principles and applications:** John Wiley & Sons; 2007.
532. Kraj A, Desiderio DM, Nibbering NM: **Mass spectrometry: instrumentation, interpretation, and applications,** vol. 20: John Wiley & Sons; 2008.
533. Chen C-HW: **Review of a current role of mass spectrometry for proteome research.** *Analytica chimica acta* 2008, **624**(1):16-36.
534. Herbert CG, Johnstone RA: **Mass spectrometry basics:** CRC press; 2002.
535. Parker CE, Warren MR, Mocanu V: **5 Mass Spectrometry for Proteomics.** *Neuroproteomics* 2009:71.

536. van Welie I, du Lac S: **Bidirectional control of BK channel open probability by CAMKII and PKC in medial vestibular nucleus neurons.** *Journal of Neurophysiology* 2011, **105**(4):1651-1659.
537. Miller C: **An overview of the potassium channel family.** *Genome biology* 2000, **1**(4):reviews0004. 0001.
538. Yuan P, Leonetti MD, Pico AR, Hsiung Y, MacKinnon R: **Structure of the human BK channel Ca²⁺-activation apparatus at 3.0 Å resolution.** *Science* 2010, **329**(5988):182-186.
539. Li G, Yan W, Dang Y, Li J, Liu C, Wang J: **The role of calcineurin signaling in microcystin-LR triggered neuronal toxicity.** *Scientific reports* 2015, **5**.
540. Moffatt BA, Ashihara H: **Purine and Pyrimidine Nucleotide Synthesis and Metabolism.** *The Arabidopsis Book / American Society of Plant Biologists* 2002, **1**:e0018.
541. Davy SK, Allemand D, Weis VM: **Cell Biology of Cnidarian-Dinoflagellate Symbiosis.** *Microbiology and Molecular Biology Reviews : MMBR* 2012, **76**(2):229-261.
542. Bellis ES, Howe DK, Denver DR: **Genome-wide polymorphism and signatures of selection in the symbiotic sea anemone Aiptasia.** *BMC Genomics* 2016, **17**:160.
543. Rocha J, Peixe L, Gomes NCM, Calado R: **Cnidarians as a Source of New Marine Bioactive Compounds-An Overview of the Last Decade and future Steps for Bioprospecting.** *Marine Drugs* 2011, **9**:26.
544. Ventola CL: **The antibiotic resistance crisis: part 1: causes and threats.** *Pharmacy and Therapeutics* 2015, **40**(4):277.
545. Tsai IH, Wang YM, Cheng AC, Starkov V, Osipov A, Nikitin I, Makarova Y, Ziganshin R, Utkin Y: **cDNA cloning, structural, and functional analyses of venom phospholipases A(2) and a Kunitz-type protease inhibitor from steppe viper Vipera ursinii renardi.** *Toxicon* 2011, **57**(2):332-341.
546. Kiyatkin N, Dulubova I, Grishin E: **Cloning and structural analysis of alpha-latroinsectotoxin cDNA. Abundance of ankyrin-like repeats.** *Eur J Biochem* 1993, **213**(1):121-127.

