THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

A combined proteomic and transcriptomic investigation of venom from Australian sea anemones

Bruno Madio

BSc Marine Biology, MSc General Physiology

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2018

Institute for Molecular Bioscience

Abstract

Cnidarians (classes Anthozoa, Scyphozoa, Cubozoa and Hydrozoa) are the oldest venomous animals on Earth, and they have evolved a diverse range of toxins to aid in prey capture, defence against predators, and intra- and inter-specific competition. More than a century of research on sea anemone venoms has shown that they contain a rich diversity of biologically active proteins and peptides. However, recent omics studies have revealed that much of the venom remains unexplored. By combining the complementary approaches of transcriptomics, proteomics, phylogenetics, mass spectrometry imaging and nuclear magnetic resonance spectroscopy this thesis provides the first holistic overview of the venom arsenal of sea anemones.

In the last decade, next-generation sequencing has become a widely used tool to investigate the diversity of components in sea anemone venom. When compared to other venomous lineages outside of Cnidaria, sea anemones are atypical venomous animals, as they express venom throughout their body instead expressed in a specific venom gland. This makes it difficult to distinguish which toxin-like transcripts are functionally venomous. The first experimental part of this thesis (Chapter 2) highlights the importance of using proteomics of milked venom to correctly identify venom proteins/peptides, both known and novel, while minimizing the number of false positive identifications from non-toxin homologues identified in transcriptomes of venom-producing cells.

The phylogenetic and molecular evolutionary histories of venoms contribute to our understanding of the different venom strategies employed across taxonomic families. Moreover, it can contribute to identification of functionally important amino acids and may aid in directing future biodiscovery efforts. Chapter 3 provides the first comprehensive insight into toxin recruitment events across a wide range of sea anemone taxa.

Sea anemones are known to produce an array of biologically active peptides with different 3D folds. Some of these are potent neuroactive peptides, acting on a diverse range of ion channels, such as voltage-gated sodium (Na_V) and potassium (K_V) channels, transient receptor potential channels (TRP) and acid-sensitive ion channels (ASICs). Chapter 4 describes the discovery, 3D structure, tissue localisation, and possible ion channel targets of a venom peptide with a newly identified 3D scaffold for sea anemone toxins.

Although there is still much to learn about the composition of sea anemone venoms and the role of individual venom components in prey capture, defence and intraspecific competition, this work provides a solid foundation for future research into the ecology and evolution of these venoms.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications during candidature

Journal Articles

Jouiaei M., Yanagihara A.A., Madio B., Nevalainen T.J., Alewood P.F., Fry B.G. (2015) **Ancient venom systems: a review on Cnidaria toxins.** *Toxins* **7**, 2251–2271

Walker A.A., Madio B., Jin J., Undheim E. A., Fry B. G., King G. F. (2017) **Melt With This Kiss: Paralyzing and Liquefying Venom of The Assassin Bug Pristhesancus plagipennis (Hemiptera: Reduviidae).** *Molecular and Cellular Proteomics* 16, 552 – 566

Madio, B., Undheim, E.A.B. and King, G.F. (2017) **Revisiting venom of the sea anemone *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus**. *Journal of Proteomics* 166:83–92

Mitchell M.L., Hamilton, B.R., Madio, B., Morales, R.A.V., Tonkin-Hill, G.Q., Papenfuss, A.T., Purcell, A.W., King, G.F., Undheim, E.A.B., and Norton, R.S. (2017) **The use of imaging mass spectrometry to study peptide toxin distribution in Australian sea anemones**. *Australian Journal of Chemistry* 70(11):1235–1237

Conferences abstracts

Madio B., Undheim E.A.B., King G.F. **Proteomic and transcriptomic investigation of the venom from Australian sea anemones provides insight into venom evolution and ecology (Poster).** 18th World Congress of the International Society for Toxinology, University of Oxford, UK, September 25–30, 2015.

Madio B., Undheim E.A.B., King G.F. **Proteomic and transcriptomic investigation of the venom from Australian sea anemones provides insight into venom evolution and ecology (Talk).** XIII Congress of the Brazilian Society of Toxinology, Campos do Jordão, Brazil, November 8–11, 2015.

Madio B., Hamilton, B. R., Chin, Y, K. Y., Dekan, Z., Alewood, P. F., Undheim E.A.B., King G.F. **A holistic investigation of the venom from Australian sea anemones.** East Coast Protein Meeting, Coffs Harbour, July 14–16, 2017.

Publications included in this thesis

Madio, B., Undheim, E.A.B. and King, G.F. (2017) **Revisiting venom of the sea anemone *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus**. *Journal of Proteomics* 166:83–92.

Contributor	**Statement of contribution**
Madio, Bruno (Candidate)	Conception and design (60%) Analysis and interpretation (70%) Drafting and production (60%)
Undheim, Eivind A.B.	Conception and design (20%) Analysis and interpretation (20%) Drafting and production (25%)
King, Glenn F.	Conception and design (20%) Analysis and interpretation (10%) Drafting and production (15%)

Contributions by others to the thesis

Prof. Glenn King (UQ) was my primary supervisor and provided significant advice on project design and coordination, in addition to editing and reviewing original research articles and this thesis. Dr Eivind A.B. Undheim acted as my cosupervisor and provided significant input into experimental design, analysis, reviewing and editing original research articles and this thesis.

Chapter 4

Pharmacological activity was done by Dr Jennifer J. Smith and Ben Cristofori-Armstrong, at IMB, UQ, Australia, and Steve Peigneur from Jan Tytgat's group at KU Leuven, Belgium. MALDI imaging mass spectrometry and histology was done under the supervision of Dr. Brett Hamilton. Ultra-high mass resolution analysis by MALDI-FT-ICR-MSI was done by Dr Berin A. Boughton at School of Bioscience, The University of Melbourne. NMR data acquisition and processing was done under the supervision of Dr Yanni K. Y. Chin. Peptide synthesis was done by Zoltan Dekan from Paul Alewood's group at IMB, UQ. Membrane interaction, hemolytic studies and cell viability assays were done by Sónia T. Henriques.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Research Involving Animal Subjects

The sea anemones were collected in Moreton National Park under the permit QS2014/MAN259 (document in the appendices section) and maintained and studied under the UQ Biosafety approval: IBC/091B/IMB/2014.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Glenn F. King for giving me the opportunity to carry out my research in his group. I would also like to thank my co-supervisor Dr Eivind Andreas Baste Undheim for all his support, friendship and who has been a source of inspiration. This great challenge to venture into venomic research would not have been so fruitful without their valuable experience and expertise. I am truly grateful.

I would also like to thank my PhD Thesis Committee members, Prof. Lachlan Rash and A/Prof. Bryan Fry, for their time and valuable comments with this project. A very special thank you to Dr. Amanda Carozzi, Ms Olga Chaourova and Ms Cody Mudgway for the immeasurable support with the application and submission process and for all of the milestone during my candidature.

Throughout my PhD I have also fostered collaborations and friendships with several amazing scientists that have helped me along the way. This project would not be possible without their collaborations. In particular, I would like to thank Alun Jones for all help with mass spectrometry and for his friendship. Also, Hareshwar Goswami for his advices in 2D SDS-PAGE electrophoresis. Dr Brett Hamilton, thanks for the long days of IMS data analyses and great music. Steve Peigneur and Prof Yan Tytgat at the The Katholieke Universiteit Leuven, Belgium many thanks for always be keen for collaborations.

Furthermore, I would like to thank all friends and colleagues in the King group as well as the IMB. Special thanks for those who helped me during field trips.

Last, but not least, I would like to thank my family for encouraging me and supporting my decision to study overseas in Australia. Even though we are so far, I know you were always there for me.

Financial support

This research was supported by the Brazilian training program scholarship ‘Science without Boarders’ and The University of Queensland Research Higher Degree Scholarship (AUOSTIP — RHD Scholarship).

Keywords

sea anemone venom, proteomics, transcriptomics, ion channel, mass spectrometry imaging, NMR, evolution

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060101, Analytical Biochemistry, 60%

ANZSRC code: 030406 Proteins and Peptides, 40%

ANZSRC code: 060399 Evolutionary Biology not elsewhere classified, 10%

Fields of Research (FoR) Classification

FoR code: 0304 Medicinal and Biomolecular Chemistry, 40%

FoR code: 0301 Analytical chemistry, 40%

FoR code: 0603, Evolutionary Biology, 20%

Table of contents

ABSTRACT I

DECLARATION BY AUTHOR III

PUBLICATIONS DURING CANDIDATURE IV

JOURNAL ARTICLES IV

CONFERENCES ABSTRACTS IV

PUBLICATIONS INCLUDED IN THIS THESIS V

CONTRIBUTIONS BY OTHERS TO THE THESIS VI

CHAPTER 4 VI

STATEMENT OF PARTS OF THE THESIS SUBMITTED TO QUALIFY FOR THE AWARD OF ANOTHER DEGREE VI

RESEARCH INVOLVING ANIMAL SUBJECTS VI

ACKNOWLEDGEMENTS VII

FINANCIAL SUPPORT VIII

KEYWORDS IX

AUSTRALIAN AND NEW ZEALAND STANDARD RESEARCH CLASSIFICATIONS (ANZSRC) IX

FIELDS OF RESEARCH (FOR) CLASSIFICATION IX

LIST OF FIGURES XIV

LIST OF TABLES XIX

LIST OF ABBREVIATIONS USED IN THE THESIS XX

CHAPTER 1 PROTEOMIC, TRANSCRIPTOMIC, AND MASS SPECTROMETRIC PROFILING OF THE ANCIENT VENOM SYSTEM OF SEA ANEMONES 1

1.1 INTRODUCTION 2

1.2 VENOM TISSUE 4

1.3 VENOM COMPOSITION **5**
1.3.1 NON-PROTEINACEOUS VENOM COMPONENTS 7
1.3.2 ENZYMES 7
1.3.3 NON-ENZYMATIC PROTEINS – CYTOTOXINS 8
1.3.4 NON-ENZYMATIC PROTEINS – NEUROTOXINS 11
1.3.4.1 ATX III 12
1.3.4.2 β-defensins 13
1.3.4.3 Boundless β-hairpin 16
1.3.4.4 EGF-like peptides 18
1.3.4.5 Inhibitor Cystine Knot fold 18
1.3.4.6 ShK motif 19
1.3.4.7 Kunitz-domain 21
1.3.4.8 SCRiPs 23
1.4 SUMMARY AND SIGNIFICANCE OF PROJECT **24**
1.5 AIMS **25**

CHAPTER 2 REVISITING VENOM OF THE SEA ANEMONE *STICHODACTYLA HADDONI*: OMICS TECHNIQUES REVEAL THE COMPLETE TOXIN ARSENAL OF A WELL-STUDIED SEA ANEMONE GENUS **26**

ABSTRACT **27**
BIOLOGICAL SIGNIFICANCE **27**
HIGHLIGHTS **27**
2.1 INTRODUCTION **28**
2.2 MATERIALS AND METHODS **29**
2.2.1 SPECIMEN AND VENOM COLLECTION 29
2.2.2 2DE ANALYSIS 29
2.2.3 HPLC 30
2.2.4 PROTEIN IDENTIFICATION USING LC-MS/MS 31
2.2.5 TRANSCRIPTOME SEQUENCING, ASSEMBLY, AND EXPRESSION ANALYSIS 32
2.2.6 FUNCTIONAL ANNOTATION OF TRANSCRIPTOME 32
2.2.7 FUNCTIONAL ANNOTATION OF PROTEOMICS DATA 33
2.3 RESULTS **33**
2.3.1 TRANSCRIPTOME SEQUENCING, ASSEMBLY AND ANNOTATION 33
2.3.2 PROTEOMICS OF MILKED VENOM 37
2.4 DISCUSSION **45**

2.5 CONCLUSIONS **48**

CHAPTER 3 EVOLUTION OF THE SEA ANEMONE VENOM ARSENAL **49**

3.1 INTRODUCTION **51**

3.2 RESULTS AND DISCUSSION **52**

3.2.1 NEUROTOXINS 58

3.2.1.1 ICK-like 61

3.2.1.2 Kunitz-like 63

3.2.1.3 β-Defensin-like 65

3.2.2 PORE-FORMING TOXINS 67

3.2.3 ENZYMES 69

3.2.4 VENOM COMPONENTS WITHOUT BLAST HOMOLOGY 70

3.3 CONCLUSION **75**

3.4 MATERIALS AND METHODS **76**

3.4.1 SPECIMEN AND VENOM COLLECTION 76

3.4.2 2DE ANALYSIS 76

3.4.3 HPLC 77

3.4.4 PROTEIN IDENTIFICATION USING LC-MS/MS 77

3.4.5 TRANSCRIPTOME SEQUENCING, ASSEMBLY, AND EXPRESSION ANALYSIS 78

3.4.6 FUNCTIONAL ANNOTATION OF TRANSCRIPTOME 78

3.4.7 FUNCTIONAL ANNOTATION OF PROTEOMICS DATA 79

3.4.8 PHYLOGENETIC ANALYSES 79

CHAPTER 4 PHAB TOXINS: A UNIQUE FAMILY OF SEA ANEMONE TOXINS EVOLVING VIA INTRA-GENE CONCERTED EVOLUTION DEFINES A NEW PEPTIDE FOLD **81**

ABSTRACT **82**

SIGNIFICANCE **82**

4.1 INTRODUCTION **83**

4.2 RESULTS **84**

4.2.1 DISCOVERY OF ATE1A 84

4.2.2 ATE1A BELONGS TO A NOVEL PEPTIDE FAMILY THAT EVOLVES BY INTRA-GENE CONCERTED EVOLUTION 86

4.2.3 ATE1A DEFINES A NEW PEPTIDE FOLD 88

4.2.4 ATE1A REPRESENTS A NEW TYPE OF SEA ANEMONE K_V-TOXIN 90

4.2.5 ATE1A IS A TOXIN WITH A PREDATORY FUNCTION 91

4.3 DISCUSSION **94**

4.3.1 ATE1A IS THE FOUNDING MEMBER OF A NEW, SIXTH TYPE OF SEA ANEMONE K_V TOXIN 94
4.3.2 ATE1A IS A PREDATORY TOXIN PRODUCED IN NEMATOCYTES 95
4.3.3 PHAB TOXINS EVOLVE VIA INTRA-GENE CONCERTED EVOLUTION 96
4.4 MATERIALS AND METHODS 97
4.4.1 VENOM COLLECTION AND FRACTIONATION 97
4.4.2 MASS SPECTROMETRY 97
4.4.2.1 MALDI-TOF/TOF Mass spectrometry 97
4.4.2.2 MALDI Mass Spectrometry Imaging 98
4.4.3 TRANSCRIPTOMICS 99
4.4.4 EVOLUTION OF ATE1A 100
4.4.5 ATE1A SYNTHESIS 100
4.4.6 ATE1A STRUCTURE 101
4.4.7 ACTIVITY OF ATE1A 101
4.4.7.1 Electrophysiological characterisation of Ate1a 101
4.4.7.2 Antimicrobial Assays 104
4.4.7.3 Interactions of Ate1a with lipid bilayers 105
4.4.7.4 Hemolysis studies 105
4.4.7.5 Cell culture and cell viability assays 105
4.4.7.6 Toxicity bioassays 106

CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS 107

5.1 SEA ANEMONE VENOM COMPOSITION 108
5.2 BIODISCOVERY POTENTIAL OF SEA ANEMONE VENOMS 109
5.3 FUTURE DIRECTIONS 110
5.4 CONCLUSIONS 110

REFERENCES 111

APPENDICES 131

COLLECTION PERMIT 131
APPENDICES OF CHAPTER 2 135
APPENDICES OF CHAPTER 3 155

List of Figures

Figure 1.1 – (A) Phylogenetic tree of Cnidarians. Representative Medusozoa depicted here are *Hydra viridis* (Hydrozoa), *Aurelia aurita* (Scyphozoa), *Chironex fleckeri* (Cubozoa) and *Haliclystus sp* (Staurozoa). The Anthozoa are the sea anemone *Nematostella vectensis* (left) and the coral *Acropora millepora* (right). Figure modified from Technau & Schwaiger (1). (B) Phylogenetic relationships among major lineages of sea anemones (After Rodrigues *et al,*, 2013 (2)). 3

Figure 1.2 – 3D structures of actinoporins as exemplified by (A) Sticholysin I (PDB accession code 2KS4) and (B) Equinatoxin II (PDB accession code 1KD6). The N- and C-termini are labelled. 10

Figure 1.3 – Alignment of representative sea anemone toxins that adopt ATXIII motif. Disulfide bridge connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown. 12

Figure 1.4 – 3D structure of the sea anemone toxin Av3 (PDB accession code 1ANS). The three disulfide bonds are represented by orange tubes. 13

Figure 1.5 – 3D structure of sea anemone toxins adopting a β-defensin structural motif. (A) . The three disulfide bonds are represented by orange tubes. Toxins are grouped according to their molecular targets. (B) Alignment of representative sea anemone toxins that contain a β-defensin motif. Disulfide connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown. 14

Figure 1.6 – Superimposed cartoon representation of the structures of CgNa (Δ-actitoxin-Cgg1a) and ApB (left) and CgNa and ShI (Δ-stichotoxin-She1a) (right). CgNa is coloured green and dark grey. Figure modified from Salceda *et al.* (119)..... 15

Figure 1.7 – 3D structure of the sea anemone toxin Ugr9-1 (PDB accession code 2LZO). The two disulfide bonds are represented by orange tubes. 17

Figure 1.8 – Alignment of representative sea anemone toxins that adopt the BBH motif. Disulfide connectivities are indicated. Amino acid identities (black boxes) and similarities (grey boxes) are shown. 18

Figure 1.9 – Alignment of sea anemone toxins that likely adopt an ICK motif. Predicted disulfide connectivities are indicated. Amino acid identities (black boxes) and similarities (grey boxes) are shown. 19

Figure 1.10 – 3D structure of K_V type 1 sea anemone toxins. (A) Structure of BgK (PDB accession code 1BGK). (B) Structure of ShK (PDB accession code 1ROO). The three disulfide bonds are represented by orange tubes. 20

Figure 1.11 – 3D structure of the K_V type 2 sea anemone toxin ShPI-I (π-stichotoxin-She2a) (PDB accession code 3OFW). Peptides of this type are homologous to Kunitz-type inhibitors of serine proteases. The three disulfide bonds are represented by orange tubes. 22

Figure 1.12 – Aligment of APHC1, 2 and 3 (UniProt B2G331, C0HJF3 and C0HJF4, respectively). Cysteines are highlighted in bold, mutations are marked with dots, and conserved positions are marked with *. 23

Figure 1.13 – Representative sea anemone SCRiP toxins. Disulfide connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown. 23

Figure 1.14 – 3D structure of the sea anemone toxin Ueq 12-1 (PDB accession code 5LAH). The three disulfide bonds are shown as orange tubes. 24

Figure 2.1 – Putative toxin families identified in the transcriptome of *S. haddoni* by BLAST search against UniProt. The 508 unique protein sequences with significant BLAST hits to the manually curated list of animal toxins in UniProt (www.uniprot.org/program/Toxins) were sorted into toxin families according to their cysteine scaffold and amino acid sequence. The pie chart shows the proportional contribution of each toxin family to the predicted venom proteome. The number of homologues identified for each protein and peptide family is shown in parentheses after the family name..... 34

Figure 2.2 – Composition of *S. haddoni* venom. (A) C_{18} RP-HPLC chromatogram of desalted crude *S. haddoni* venom. (B) 2DE gel of *S. haddoni* venom; the first dimension was isoelectric focussing (pH 3–10) followed by 12.5% SDS-PAGE. Molecular masses of standards are indicated on right of gel. Proteins identified by in-gel digestion and LC-MS/MS are annotated according to their respective family. Spots with less than two peptides identified in the LC-MS/MS and confidence values below 95% were not annotated. 39

Figure 2.3 – Diversity of putative toxins detected in, or absent from, the venom of *S. haddoni*. Histogram of unique protein sequences within each toxin family identified in the

transcriptome that were (black) or were not (grey) identified in proteomic analyses of milked venom. Putative toxin families U_1-Std to U_{12}-Std are new scaffolds as described in Table 2.2. 43

Figure 2.4 – Transcript expression levels for putative toxins detected in, or absent from, *S. haddoni* venom. Expression levels ('transcripts per million') are shown as box plots for each toxin family. 44

Figure 3.1 — Proteomic analyses of five sea anemone venoms. Composition of venom obtained by electrostimulation is show as rpHPLC chromatrogams (left) and 2D-PAGE gels (right) for A) *Aiptasia pulchella*, B) *Actinia tenebrosa*, C) *Heteractis malu*, and D) *Macrodactyla doorensis*, as well as E) *Stichodactila haddoni* for comparison (see Chapter 2). For 2DE gels the Isoelectric point (pI) is indicated above and molecular weight (kDa) on the right of each gel. 54

Figure 3.2 – Putative broad functional venom profiles of sea anemones. Putative toxins identified in transcriptomes of sea anemones based on our proteomics database. Putative neurotoxins are cysteine-rich venom peptides related to known neurotoxic sea anemone. Total number of homologues identified is shown. 56

Figure 3.3 — Venom neurotoxin scaffold (family) diversity across sea anemones. Function annotation of the unique CDSs identified by BLAST search using our proteomics database. Identical sequences were removed using CD-HIT. Species were abbreviated according to Oliveira and colleagues (33). Each species is indicated by a number and colour as shown in the key at right side. 58

Figure 3.4 — Phylogenetic distribution of neurotoxins families in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of neurotoxin families as their earliest respective recruitments. 60

Figure 3.5 — Phylogenetic reconstruction of ICK homologues in sea anemone venoms. Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+I+G4 model, which was the best fit model according to Bayesian information criterion (BIC). 62

Figure 3.6 — Phylogenetic reconstruction of Kunitz-like peptides in sea anemone venoms.

Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+R5 model, which was the best fit model according to BIC..... 64

Figure 3.7 — Phylogenetic reconstruction of β-Defensin-like peptides in sea anemone venoms. Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+R5 model, which was the best fit model according to BIC..... 66

Figure 3.8 — Phylogenetic distribution of actinoporin and enzymes families in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of actinoporin and enzyme families as their earliest respective recruitments. 68

Figure 3.9 — Venom pore-forming toxin and enzymes (family) diversity across sea anemones. 69

Figure 3.10 — Diversity of proteins with no blast homology across sea anemones. 71

Figure 3.11 — Phylogenetic distribution of proteins with no blast homologies in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of these novel families. A Families U1 to U10, and in B Families from U11 to U20..... 72

Figure 3.12 — Representative sequence of proteins with no blast homology. Alignment of representative sequences of U1 to U10..... 73

Figure 3.13 — Representative sequence of proteins with no blast homology. Alignment of representative sequences of U11 to U20..... 74

Figure 4.1 – Isolation and sequencing of Ate1a. (A) C18 RP-HPLC chromatogram showing fractionation of crude A. tenebrosa venom. The early-eluting peak containing Ate1a is highlighted in red. Inset shows average mass and isotope family for Ate1a. (B) De novo sequencing of Ate1a using ISD-MALDI MS. 85

Figure 4.2 – Domain architecture and evolution of Ate1a precursors. (A) Domain architecture of Ate1a and Ate1a-like contigs. Prepropeptides are composed of a signal peptide (SignalP), one or two cysteine-containing propeptide domains (CysProP), and three

cysteine-free propeptide domains (LinearProP) that each precede an Ate1a-like PHAB domain. (B) Nucleotide sequence alignments for each domain. (C) Maximum likelihood phylogenetic reconstructions for each domain. Bootstrap support values are shown at the nodes, while horizontal bars indicate genetic distance. Sequence accessions are for *S. haddoni* (1–3) TR75252_c0_g2_i1–3 and *A. viridis* (1) FK754894, (2) FK726055, and (3) FK733314. 87

Figure 4.3 – 3D structure of Ate1a. (A) Solution structure of Ate1a (ensemble of 20 structures; PDB code 6AZA). Disulfide bonds are highlighted in orange and proline side chains are shown in blue. (B) Surface representation of Ate1a with cationic and uncharged residues shown in blue and grey, respectively. (C) Surface representation of Ate1a showing relative hydrophobicity, which increases from white to red. 89

Figure 4.4 – Ate1a is the first member of the new PHAB fold. Comparison of the PHAB fold with other peptide folds containing two disulfide bonds and a similar number of residues (16-29 residues). Disulfide bonds are shown as orange tubes and N- and C-termini are labelled. (A) Ate1a; (B) β-hairpin fold represented by the spider peptide gomesin (PDB 1KFP); (C) CS α/α motif represented by the scorpion toxin κ-hefutoxin1 (PDB 1HP9); (D) Boundless β-hairpin motif represented by sea anemone toxin π-AnmTX Ugr 9a-1 (PDB: 2LZO); (E) Disulfide-directed hairpin represented by scorpion toxin U_1-Liotoxin-Lw1a (PDB 2KYJ); (F) Unstructured two-disulfide peptide fold represented by centipede toxin RhTx (PDB 2MVA). 90

Figure 4.5 – Tissue distribution of Ate1a determined using MSI. (A) MSI linear positive mode spectra acquired from a cross-sectioned animal, with peaks corresponding to Ate1a filled in. The spectrum of native tissue is blue while the spectrum obtained after on-tissue gas-phase reduction and alkylation is shown in red. Inset shows a mass difference of 180 Da, corresponding to ethanolylation of four cysteine residues. (B) Ultra-high mass resolution analysis of the peak corresponding to Ate1a acquired by MALDI-FT-ICR-MSI at a resolution of 16,000,000 and resolving power at 1890 *m/z* of > 500,000, showing observed (top) and calculated (bottom) spectra. (C) Left: Histological image of the sea anemone section used for MSI experiments, stained with hematoxylin and eosin. Right: distribution of the peak corresponding to the average mass of Ate1a as observed by MALDI-TOF MSI. 93

List of Tables

Table 2.1 – Highly expressed toxin transcripts. Ten most highly expressed transcripts encoding putative toxins in the tentacle transcriptome of *S. haddoni* as identified by BLAST search against the manually curated list of animal toxins in UniProt (www.uniprot.org/program/Toxins). Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit. 37

Table 2.2 – Novel toxin scaffolds identified in *S. haddoni* venom. Novel protein and peptide families identified in the venom of S. *haddoni* organized into families according to cysteine scaffold and amino acid sequence similarity. 'X' indicates any amino acid. 40

Table 2.3 – Most highly expressed transcripts identified in milked venom. Ten most highly expressed toxin candidates identified from proteomic analysis of *S. haddoni* venom. Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit. 41

Table 2.4 – Toxin homologues not detected in the venom proteome. Ten most highly expressed putative toxins with *homology* to sequences identified in the venom of *S. haddoni*. Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit. 44

Table 3.1 — Summary of protein families in proteomics of sea anemone venoms. Table shows in which species in this study was identified the toxin families listed. 55

List of Abbreviations used in the thesis

1,5-DAN	1,5-Diaminonaphthalene
2D-PAGE	2-Dimensional Polyacrylamide Gel Electrophoresis
2DE	2 Dimensional Electrophoresis
5-HT	5-hydroxytryptamine
ACN	Acetonitrile
AMP	Antimicrobial peptide
ASIC	Acid-sensing ion channel
BBH	Boundless β-hairpin
BDS	Blood Depressing Substances
BLAST	Basic Logical Alignment Search Tool
BSA	Bovine serum albumin
CAP	CRiSP (cysteine rich proteins), Allergen-5, and Pathogenesis-related
Ca_V	Voltage-gated Calcium Channel
cDNA	Complementary DNA
CDS	Coding sequence
CHCA	α-Cyano-4-hydroxycinnamic Acid
CID	Collision-Induced Dissociation
cRNA	Complementary RNA
CSαα	Cystine-stabilised α/α

DDH	Disulfide-directed Hairpin
DIEA	Diisopropylethylamine
DMF	Dimethylformamide
EGF	Epidermal Growth Factor
ESI-MS	Electro-Spray Ionization Mass Spectrometry
EST	Expressed Sequence Tag
Et2O	Diethyl Ether
FA	Formic Acid
FDR	False Discovery Rate
HCD	Higher-energy Collisional Dissociation
HCTU	O-(1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEK293	Human Embryonic Kidney 293
hERG	Ether a go-go related gene
HSQC	Heteronuclear Single Quantum Coherence
ICK	Inhibitor Cystine-knot
IEF	Isoelectric Focusing
IPG	Immobilized pH Gradient
ISD	in-source Dissociation
K_{Ca}	Calcium activated potassium channel
kDa	kilo Dalton

KV	Voltage-gated potassium channel
LC	Liquid Chromatography
m/z	Mass to charge ratio
MACPF	membrane-attack complex/perforin
MALDI-FT-ICR-MSI	Matrix-Assisted Laser Desorpion/Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Imaging
MHB	Mueller Hinton Broth
MIC	Minimal inhibitory concentration
mRNA	Messenger RNA
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
Mya	Million Years Ago
Na_V	Voltage-gated sodium channel
NBS	Non-binding Surface
NL	Non-linear
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
PDB	Protein Data Bank
PFTs	pore forming toxins
PHAB	Proline-Hinged Asymmetric b-hairpin
PLA_2	Phospholipase A_2

POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine
ROI	Region-of-interest
RP-HPLC	Reverse-Phase High-Performance Liquid Chromatography
SCRiPs	Small Cysteine-rich Peptides
SPR	Surface Plasmon Resonance
SUV	Small Unilamellar Vesicle
TFA	Trifluoroacetic Acid
TOCSY	Total correlation spectroscopy
TOF	Time of Flight
TPM	Transcripts Per Million
TRP	Transient Receptor Potential
TRPA1	Transient Receptor Potential Ankyrin-repeat 1
TRPV1	Transient Receptor Potential cation channel subfamily V member 1
V-h	Volt-hours
YNB	Yeast Nitrogen Broth
α-PFTs	α-pore-forming toxins

CHAPTER 1
Proteomic, transcriptomic, and mass spectrometric profiling of the ancient venom system of sea anemones

1.1 Introduction

Sea anemones, though known as the flowers of the sea, are exclusively marine animals that belong to the phylum Cnidaria (Figure 1.1A). Essentially laminar organisms, their two-dimensional epithelial construction has shaped both behavioral and physiological responses and led to great ecological success despite their structural simplicity, as evidenced by their presence in all marine ecosystems. They also perform a key role in benthic–pelagic coupling as part of the benthic suspension feeding community (3), transferring energy to the benthos from the water column and releasing metabolites, gametes, and offspring back into the water column.

Sea anemones belong to the class Anthozoa, which differ from all other cnidarians in that they lack a free-swimming medusa stage. Within Anthozoa, sea anemones form the hexacorallian order Actiniaria, which contains only solitary, sessile, benthic polyps. There are around 1200 species of sea anemones organized in 46 families and they constitute the greatest diversity within Anthozoa. The life cycle of sea anemones comprises sexual reproduction and an asexually budding phase. Polyps may be single sex or both male and female, and the sexual life cycle is straightforward including four main stages: the fertilized egg, planula larvae, polyp and sessile sea anemone. Interestingly, sea anemones have great powers of regeneration (4) and can reproduce asexually in multiple ways: by budding, fragmentation, or by longitudinal or transverse binary fission (5).

Relationships within Actiniaria as determined by phylogenetic analyses of DNA or morphological characters do not accord with the divisions of the traditional classification, and the order was consequently recently revised so that taxonomic divisions correspond to phylogenetic relationships (6). The primary division within the order is between the Anenthemonae and Enthemonae. Anenthemonae is the less species-rich of the suborders, containing members of families Actinernidae, Edwardsiidae, and Halcuriidae; the model organism *Nematostella vectensis* is the most familiar and well-studied member of this group. Enthemonae contains the overwhelming majority of species and anatomical diversity within Actiniaria and it is further subdivided into the superfamilies Actinioidea, Actinostoloidea, and Metridiodea (Figure 1.1B).

Sea anemones are commonly considered a group of exclusively predatory animals, however they are also opportunistic, omnivorous suspension feeders. The dietary composition of species varies markedly between different marine habitats, reflecting the different composition of the

macrobenthic organismic assemblages in different areas (7). Sea anemones capture prey that come within reach of their tentacles, enabling them to immobilize the prey with their venom. The mouth can stretch to help in prey capture and ingestion of larger animals such as crabs, molluscs and even fish (8). However, some sea anemones also feed on organic detritus, which are caught with the aid of a mucus secretion. In addition, many sea anemones form a facultative symbiotic relationship with zooxanthellae, zoochlorellae or both. These single-celled algal species reside in the anemone's gastrodermal cells, especially in the tentacles and oral disc. The sea anemone benefits from the products of the algae's photosynthesis and the algae in turn are assured protection and exposure to sunlight (9).

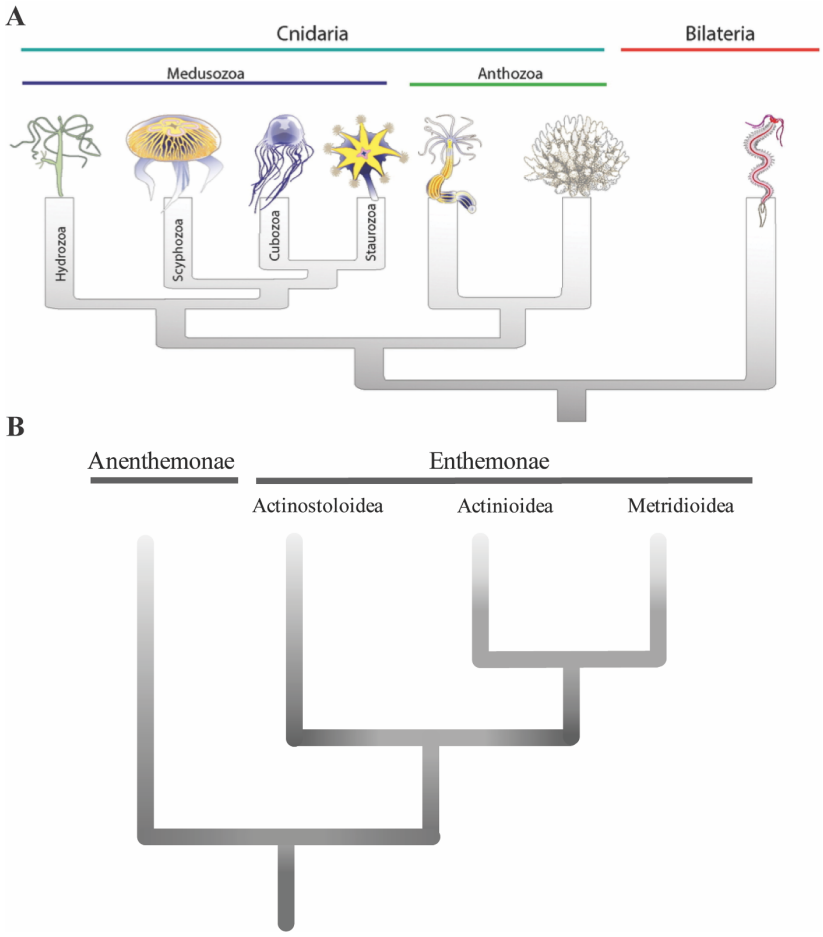


Figure 1.1 – (A) Phylogenetic tree of Cnidarians. Representative Medusozoa depicted here are *Hydra viridis* (Hydrozoa), *Aurelia aurita* (Scyphozoa), *Chironex fleckeri* (Cubozoa) and *Haliclystus sp* (Staurozoa). The Anthozoa are the sea anemone *Nematostella vectensis* (left) and the coral *Acropora millepora* (right). Figure modified from Technau & Schwaiger (1)**. (B)** Phylogenetic relationships among major lineages of sea anemones (After Rodrigues *et al,*, 2013 (2)).

1.2 Venom tissue

Cnidarians represent the only lineage of venomous animals that lack a centralized venom system. Instead of a venom gland, sea anemones produce venom in tissues throughout the body using two different type of cells, known as nematocytes and ectodermal gland cells (10, 11). Nematocytes, which are present in all cnidarians, produce highly complex venom-filled organelles known as nematocysts. Nematocysts are the primary venom delivery apparatus of cnidarians, and they are made of a capsule containing an inverted tubule capable of extremely fast and powerful discharge (12, 13). There are at least 25 different types of nematocysts found in sea anemones, with multiple types harboured by a single specimen (14). Moreover, distinct morphological regions of a sea anemones have specialized structures and they are defined by a specialized complement of nematocysts (15). Examples of functional specialization of the venom in different tissues includes tentacles used for prey capture, immobilisation and defence; acrorhagi for competition and defence; column for external defence; and actinopharnyx and mesenterial filaments, both used in prey immobilisation and digestion (16). The ecological and evolutionary success of cnidarians since the Cambrian explosion may be to a large degree attributed to this complex organelle system and the toxins it contains.

In addition to nematocytes, sea anemones also produce toxins in a second type of cell known as an ectodermal gland cell, which may or may not produce distinct repertoires of toxins compared to nematocysts (10, 17). The reason why sea anemones have their toxins located in two different types of cells remains unknown. However, secretion by gland cells may allow for delivery of larger amounts of the toxin, and present an opportunity to extend the reach of venom use by the anemone beyond direct contact. So far, gland cells have only been reported to be present in *Anthopleura elegantissima*, *Anemonia viridis* and *Nematostella vectensis* (18). Thus, whether venom-secreting ectodermal gland cells are an adaptation to the environment and diet of some anemones, or whether they represent the ancestral venom-secreting cell type, remains to be determined through investigation of additional sea anemones species.

In general, most sea anemones are relatively harmless to humans. However, the venom of some species can cause severe effects. The envenomation capabilities may be linked to size and types of nematocysts. For example, the extremely large basitrich nematocysts found in the balloon-like extensions of branching tentacles (acrospheres) of some sea anemones (19) may be capable of penetrating the epidermis, explaining the severe symptoms observed in humans (20, 21).

1.3 Venom composition

Although sea anemones are flexible in the ways in which they obtain nutrition (22), they are fundamentally predatory animals, using their tentacles to catch prey. Because they lack true muscle tissue, have no visual capacity, and lack a centralized or coordinated nervous system, prey capture relies heavily on toxins to subdue prey. Like many venomous lineages, characterization of toxic components in sea anemones has been done mostly through an opportunistic approach, focusing on peptides and taxa that are easily accessible and that potentially have therapeutic relevance. As a result, the venom composition in most species remains unknown despite decades of research (23, 24). Nevertheless, sea anemone venoms have been shown to be complex mixtures of proteins, peptides and non-proteinaceous compounds. The main components found in the sea anemone venom are traditionally grouped into four functional types: (1) Phospholipase A_2 that degrades membrane phospholipids of neuronal and muscle cells, causing nerve damage and muscle inflammation (25); (2) Cytolysins that act on cell membranes and cause cell lysis (26); (3) Neurotoxins that interact with voltage-gated ion channels, acid-sensing ion channels (ASICs), and transient receptor potential ion (TRP) channels (23, 27-29), thereby altering neural transmission (30, 31); (4) Non-proteinaceous compounds (e.g. purines, biogenic amines) that are believed to induce pain during envenomation.

Until recently, no systematic nomenclature existed for naming and organising sea anemone toxins. This resulted in multiple names being assigned to the same toxin, toxins from unrelated species being designated by the same name, and ambiguous name designations. However, in 2012 two articles were published that suggested a rational nomenclature for naming sea anemones toxins. Kozlov and Grishin (32) suggested a nomenclature for cysteine-rich polypeptides toxins from sea anemones, while Oliveira and colleagues (33) suggested a more general nomenclature for naming any kind of sea anemones toxin. Moreover, Norton (34) suggested that in order to avoid confusion with sea anemone toxins and other anemone venom peptides, the nomenclature could be modified to include toxin type, such as Types K1, K2, and K3 for potassium channel toxins.

Here, the criteria stipulated by Oliveira are followed, because they can be used for all types of toxins, the species can be identified, and it avoids confusions with similar names. This sea anemone toxin nomenclature is similar to that previously proposed for spider toxins by King and colleagues (35) and adapted by Undheim and colleagues (36) for naming centipede toxins. This nomenclature consists of five terms: The first term is a Greek symbol that serves as a

broad activity descriptor denoting the molecular target. The second term is a generic name indicating the taxonomic family. The third term is a three-letter code signifying the species of origin, and consists of an initial uppercase corresponding to the first letter of the genus, followed by two lowercase letters that indicate species. The two lasts terms are formed by an alphanumeric descriptor to assign the chronological order of sequence deposition into public database or original publication related to the toxin. The alphabetic character indicates the paralogous relationship which is assigned based on amino acid sequence analysis. Within the same species, toxins sharing a high level of sequence identity and similarity are clustered in an isoform or isotoxin group. In each isotoxin group, lowercase Latin letters are also given in alphabetic order according to the date of the toxin sequence published. The nomenclature is not italicized, has hyphens to separate the first three terms, and relies on the taxonomic sea anemone names included in the Hexacorallians of the Word database (http://geoportal.kgs.ku.edu/hexacoral/ anemone2/index.cfm) to generate a three-letter code for designating sea anemone species.

Sea anemone toxins have traditionally been classified according to their activity, similarity of amino acid sequence, and the pattern of disulfide bridges (number and distribution of cysteines) (30). In this way, the toxins are first divided by their molecular target (e.g. Na_V, K_V, etc.) and then by types according to their similarity and mechanism of action. However, it is known that proteins and peptides with certain structural characteristics have been more often recruited into venoms and subsequently undergone functional radiation (37). In most cases, the stabilization of these molecular scaffolds through disulfide bonds facilitates modifications of non-structural residues, allowing alterations of surface-exposed residues without affecting the structural core (38). This means that the current classification system suffers from vulnerability to functional convergence (e.g. toxins with Kunitz or defensins scaffolds both interact with K_V channels) as well as functional promiscuity (e.g. APETx2 interacts with both ASIC and K_V channels). In addition, for most known sea anemone toxins, the exact receptors they target (e.g. ion channel subtypes) is either unknown, or at best incomplete. Because of this, the following discussion on the main components of sea anemone venoms is not divided into the traditional toxin types. Instead distinctions are made firstly between proteins and non-proteinaceous compounds, secondly between enzymes and other proteins without enzymatic activity, then according to the structural scaffold, and finally according to molecular targets.

1.3.1 Non-proteinaceous venom components

Sea anemones are known to be a rich source of protein and peptide toxins. In contrast, little is known about the non-peptidic components of their venoms. The first small molecule described from a sea anemone was a purine derivative isolated in 1986 from the Brazilian sea anemone *Bunodosoma caissarum* and named caissarone (39). Caissarone induced twitching in electrically stimulated guinea pig ileum-myenteric plexus, and this was attributed to antagonistic actions on the adenosine receptor (40). Years later, several groups began investigating small-molecule fractions based on a study reporting the antagonism of glutamate receptors by a low molecular weight fraction from venom of the Caribbean sea anemone *Phyllactis flosculifera* (41). As result of these studies, an acylated amino acid, bunodosine 391, was isolated from venom of the Brazilian sea anemone *Bunodosoma cangicum*, which likely acts on 5-hydroxytryptamine (5-HT; serotonin) receptors. Bunodosine 391 was analgesic in animal models of pain, and this activity was completely blocked by methysergide, a nonselective 5-HT receptor antagonist, but not the opioid receptor antagonist naloxone. Although this is an interesting topic with therapeutic relevance, nothing further has been published, and small molecules from sea anemone venoms remain unexplored.

1.3.2 Enzymes

There is a large discrepancy between the types of enzymes reported from transcriptomic studies of sea anemones with those reported from studies on milked venom. As previously mentioned, sea anemones do not have a centralised venom gland, and it is therefore difficult using transcriptomic techniques to distinguish between enzymes that have housekeeping roles and those that play a role in envenomation. Because of this uncertainty, only enzymes purified from sea anemone venom will be described in this chapter.

To date, PLA_2 is the only enzyme identified in the milked venom of sea anemones. PLA_2 catalyses the hydrolysis of phospholipids into free fatty acids and lysophosholipids. As phospholipids are one of the main chemical constituents of the cell envelope, enzymes capable of hydrolyzing these molecules, such as PLA_2, are likely to cause membrane disruption. The PLA_2 superfamily currently contains 15 separate groups and numerous subgroups of PLA_2 (42). In addition to sea anemones, PLA_2s are present in the venom of a wide range of venomous taxa, including snakes, bees, wasps, lizards, and centipedes (43). All these compounds belong to the group of secreted low molecular weight PLA_2s.

Sea anemone PLA_2s are poorly studied. So far, sequences have been reported for only PLA_2s isolated sea anemone venom. However, several studies have shown that sea anemone extracts (whole animal, tentacles, acontia) have PLA_2 activity. Purification and sequencing of these proteins will be important to classify and understand the ecological importance of PLA_2 in sea anemone venom.

1.3.3 Non-enzymatic proteins – Cytotoxins

Cytolysins are a group of sea anemone toxins that form pores in cell membranes, and therefore belong to a larger group of 'pore forming toxins' (PFTs) (44). The formation of pores by PFTs can be achieved by inserting either α-helices or β-hairpins within the cell membrane. Moreover, these proteins show a dual behaviour at the water-membrane interface. In water, they remain mostly monomeric and stably folded but, when they interact with lipid membranes of specific composition, they become oligomeric integral membranes structures. So far, cytolysins from more than 32 different sea anemones species have been isolated. Based on their primary structure and functional properties, cytolysins have been classified into at least four polypeptide types (types I–IV).

Type I cytolysins are 5–8 kDa peptides that form pores in phosphatidylcholine membranes, and additionally have antihistamine activity. So far, this type has been reported in just a few species, such as *Tealia felina* (accepted name *Urticina feline*) (45), and *Heteractis crispa* (previously known as *Radianthus macrodactylus*) (46). RmI (5.1 kDa) and RmII (6.1 kDa) are two cytolysins from *H. crispa* that have been biochemically characterized to some extent. They are basic peptides with pI values of about 9.2, and, in contrast to 20 kDa cytolysins, they contain cysteine residues and lack tryptophan. The structural features of this group of cytolysins are unknown.

Type II cytolysins are the most abundant and best studied cytolysins from sea anemone, and they consist of the family of the α-pore-forming toxins (α-PFTs) (47). These toxins are also known as actinoporins due to their ability to bind membrane phospholipids domains of the host organism, then oligomerize to form cation selective pores (48). Actinoporins are comprised of a single domain (~20 kDa), they lack cysteine residues, and they are equipped with functionally important regions conserved throughout the toxin gene family (49, 50). Moreover, they appear as multigene families (51), resulting in many protein isoforms. Although they display high levels of sequence identity (60–80%), the sequence differences are sufficient to cause large

differences in solubility and hemolytic activity (52).

Actinoporins have not been identified in nematocysts (53, 54), although their genetic architecture, which includes signal and propeptide motifs, indicate that they are likely synthesized in the Golgi apparatus during nematocyst development (55). Tissue-specific studies revealed that actinoporins are expressed in mesenterial filaments, suggesting that they play a role in digestion (56).

Despite most prey or predator species having a relatively conserved and ubiquitous actinoporin target site (sphingomyelin), many highly conserved actinoporins exhibit variable rates of cytolytic activities (57). This suggests that minor changes in amino acids across actinoporins may have co-diversified to target cell membranes in specific lineages (26, 55). Several residues have been manipulated to identify functionally important regions within actinoporins (58), revealing an aromatic-rich region that forms the phosphocholine (POC) binding site, with a single amino acid residue (W112 in Equinatoxin II (EqII or Δ-actitoxin-Aeq1a)) playing a key role in initiating sphingomyelin recognition and pore formation (58, 59). Although events leading to oligomerization remain uncertain, both the RGD domain (R144, G145, and D146 in EqII) and a single valine residue (V60 in EqII) are thought to direct protein attachment and play a key role in this process (60). Finally, a key arginine residue (R31 in EqII) and various hydrophobic residues in the α-helix at the N-terminal region are involved in cell membrane penetration and the formation of the ion conduction pathway (discussed below) (47).

3D structures have been reported for only three sea anemone cytolysins (Figure 1.2) and all of them are type II cytolysins (actioporins). Basically, their structure is composed of a tightly folded β-sandwich core flanked on two sides by α-helices (Figure 1.2). The first 30 residues encompass one of the helices. This is the only part of the molecule able to undergo a conformational change without any structural modification of the β-sandwich. A prominent patch of aromatic amino acids is located on the bottom of the molecule (61, 62).

Actinoporins are interesting toxins to study the ability that PFTs have to become integral membrane proteins, once they are relatively small and cysteineless. The molecular details of the mechanism by which the actinoporins so potently form pores in target membranes remain elusive. However, based on many functional studies and the structures of EqtII and StII, it is proposed that following steps occur during pore formation by the actinoporins: First the toxin attaches to the membrane via specific recognition of sphingomyelin using the aromatic-rich

region and adjacent POC binding site (59, 61-63); second, the N-terminal segment is transferred to the lipid–water interface (62, 64, 65); finally, the toxin oligomerises on the surface of the membrane and α-helices from 3 or 4 monomers insert into the membrane and form a cation-selective conduction pathway of diameter 1–2 nm (64, 66-68). This last step includes an important contribution by membrane lipids and the monomers are likely arranged in a so-called toroidal pore arrangement (69, 70).

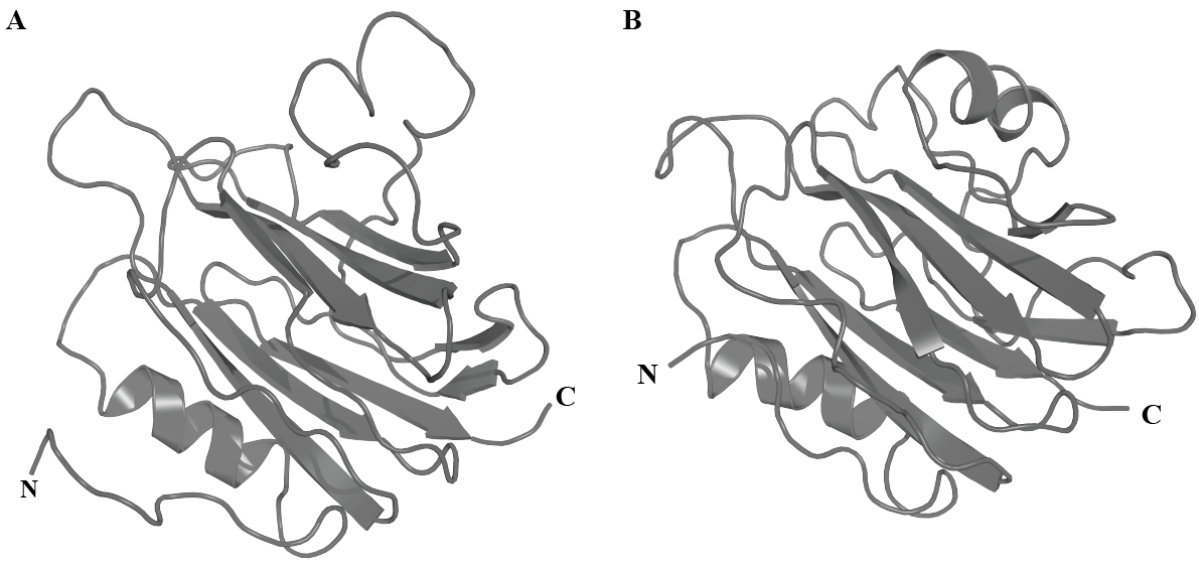


Figure 1.2 – 3D structures of actinoporins as exemplified by (A) Sticholysin I (PDB accession code 2KS4) and (B) Equinatoxin II (PDB accession code 1KD6). The N- and C-termini are labelled.

Actinoporins have been used to elucidate cell membrane dynamics, and their potential for biomedical applications has been explored (71, 72). One possible usage could be as immunotoxins (ITs), since they are extremely cytotoxic and cytolytic to a variety of cells and their vesicular organelles (73, 74). ITs are chimeric molecules in which a cell binding ligand, such as a monoclonal antibody or a growth factor are coupled to a killer toxin (in this case actinoporins) in order to address its activity towards a specific cell (for a review see Refs (75, 76)).

Type III cytolysins are 25–45 kDa proteins with or without PLA_2 activity. They were first detected in the venom of *Aiptasia pallida* (77) and later in venom of sea anemones from the genus Urticina (*U. crassicornis* and *U. piscivora*) (78, 79). In contrast with the actinoporins, these cytolysins contain several cysteine residues (78). They cause hemolysis of rat, guinea

pig, dog, pig and human red blood cells at concentrations as low as 10^{-10} M. Interestingly, the hemolytic activity of Type III cytolysins is inhibited by sphingomyelin but not by cholesterol, which is not common for actinoporins (80).

Type IV cytolysins were isolated from sea anemone homogenate as a cholesterol-inhibitable cytolysin (81). Metridiolysin from *Metridium senile* is so far the only representative of this group (26). It has a molecular mass of 80 kDa and, similarly to a group of bacterial toxins, it is activated by thiols to produce ring structures on membranes (82). Metridiolysin binds nonspecifically to lipid membranes (26), and it forms fluctuating K^+-permeable pores in planar lipid membranes (26).

Type V cytolysins are similar to the membrane-attack complex/perforin (MACPF) family of proteins. They were first discovered in nematocysts of the stinging sea anemone *Phyllodiscus semoni*. MACPF family proteins were originally identified as pore-forming factors utilised in the mammalian host defence immune system (83, 84). Similar to perforins, Type V cytolysins have an EGF-like domain close to the MACPF domain, but lack the C2 domain for attachment to lipid membranes.

Type V cytolysins were the first MACPF proteins found in non-mammalian species, and the first reported case of MACPF proteins recruited into venom. Sea anemone MACPF-cytolysins have a mass of ~60 kDa. Thus far, only three have been described (AvTx-60A from *Actineria villosa*, PsTx-60A and B from *Phyllodiscus semoni*), but they are predicted to be present in *Nematostella vectensis* based on the genome sequence of this anemone (85). The discovery and characterization of MACPF-cytolysins in sea anemones has aided our understanding of the mechanism of membrane permeabilization by MACPF proteins as well as the evolution of MACPF superfamily.

1.3.4 Non-enzymatic proteins – Neurotoxins

Neurotoxins are toxins that interfere with the transmission of nerve impulses by modifying the function of ion channels in nerve or muscle cells (86). Diverse venomous animals have evolved neurotoxins that interact with ion channels to immobilise prey and/or deter predators. Because sea anemones are sessile animals, venom neurotoxins play a critical role in immobilisation of prey and defence against predators. Neurotoxins are among the best characterized components of sea anemone venoms in terms of their mechanisms of action. They interact with a wide range of ion channels, including ASICs (24, 87, 88), TRP channels (23, 27-29), voltage-gated sodium (Na_V) channels (89-92) and voltage-gated channels

potassium (K_V) channels (93-95). Of these, K_V toxins comprise 136 of the 320 sea anemone toxins reported in UniProtKB to date (96) and thereby constitute the most diverse group, with Na_V toxins in close second place. The diversity of structural scaffolds among sea anemone neurotoxin is also remarkable. To date, eight unique structural folds have been identified based on 3D structure and/or cysteine-pattern: ATX-III, β-defensin-like, boundless β-hairpin, EGF-like, inhibitor cystine-knot (ICK), Kunitz-domain, small cysteine-rich peptides (SCRiPs), and ShK.

1.3.4.1 ATX III

ATX III toxins form a compact structural motif composed of 27–32 residues that consists of only turn-based secondary-structure elements reticulated by three disulfide bonds (Figure 1.3). These toxins lack both α-helices and β-strands (Figure 1.4) (31, 97). Several residues located on the surface of the molecule form a hydrophobic patch that may constitute part of the Na_V channel binding surface (97). To date, only six ATX III toxins have been identified from three species (*Anemonia* viridis, *Dofleinia armata* and *Entacmaea quadricolor*), and all of them target Na_V channels, delaying the process of channel inactivation. These toxins are inactive against mice (98, 99), but highly active on insects and crustaceans (100).

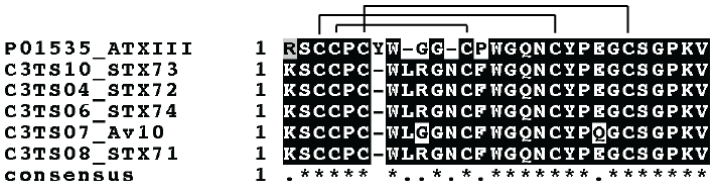


Figure 1.3 – Alignment of representative sea anemone toxins that adopt ATXIII motif. Disulfide bridge connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown.

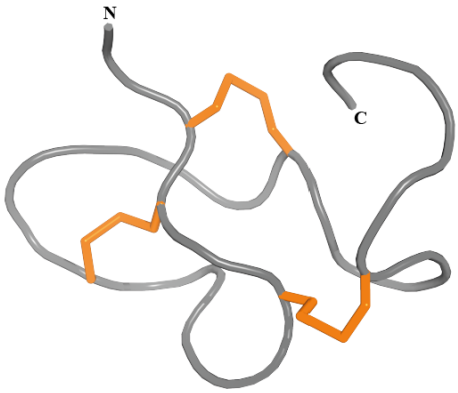


Figure 1.4 – 3D structure of the sea anemone toxin Av3 (PDB accession code 1ANS). The three disulfide bonds are represented by orange tubes.

1.3.4.2 β-defensins

The β-defensin-fold generally consists of a short helix or turn followed by a small twisted anti-parallel β-sheet (Figure 1.5). The six cysteine residues, which are paired in a 1–5, 2–4, 3–6 fashion, are crucial for maintaining the compact core configuration of β-defensins. The last two cysteines are consecutively situated (in a CCX_n pattern where $n \geq 1$) near the C-terminus. β-Defensins are antimicrobial peptides that are secreted as part of the innate immune response in a wide range of taxa (101, 102). However, in sea anemone venoms, β-defensin-like peptides have become weaponized to serve as neurotoxins that modify the activity of voltage- and ligand-gated ion channels; this family of peptides includes K_V type 3, Na_V type 1, 2 and 4 and ASIC toxins (103-106) (Figure 1.5).

K_V type 3 sea anemone toxins are composed of peptides of 42–43 residues. APETx1 (κ-actitoxin-Ael2a) from *Anthopleura elegantissima* (107) and BDS-I (Δκ-actitoxin-Avd4a) and BDS-II (Δκ-actitoxin-Avd4b) from *Anemonia sulcata* (108) are representative of this type. Although they share the same structural motif and have 40% sequence identity with BDS toxins, APETx1 has a different activity on K_V channels. Despite their classification as K_V channel toxins, it was recently shown that some K_V type 3 toxins also interact with Na_V channels (109, 110). The name BSD is an abbreviation of "Blood Depressing Substances", because these toxins were first characterized as antihypertensive and antiviral compounds (111). BDS-I and BDS-II have 93% sequence identity, and they both block the subtypes $K_V3.1$, $K_V3.2$ and $K_V3.4$ at nanomolar concentration (108, 112). Both toxins also induce a positive shift of the activation curve of $K_V3.1$ and $K_V3.2$ currents. At high concentrations they also

weakly block K_V1.1-5, K_V2.1–2, K_V4.1 and K_V4.3 currents by 2–20% (93). BDS-I is capable of modulating Na_V channel gating in a manner similar to previously known neurotoxin receptor site 3 anemone toxins, but with different isoform sensitivity (109).

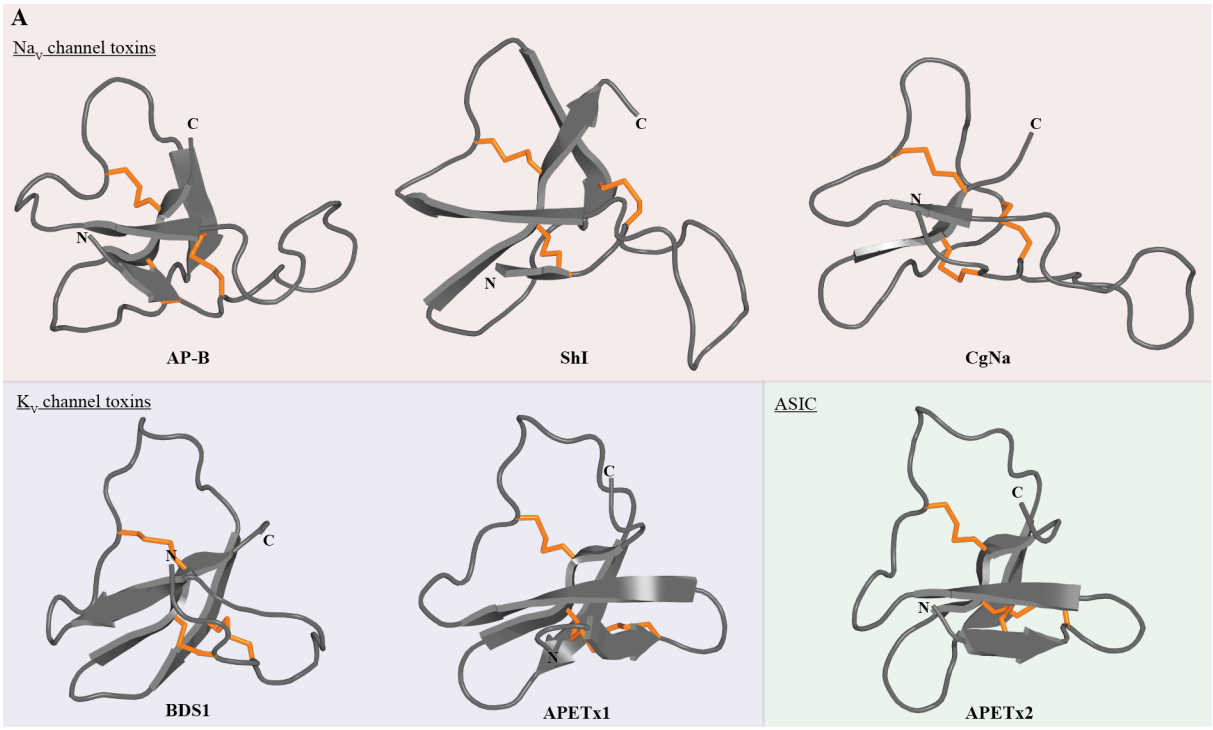


Figure 1.5 – 3D structure of sea anemone toxins adopting a β-defensin structural motif. (A) . The three disulfide bonds are represented by orange tubes. Toxins are grouped according to their molecular targets. **(B)** Alignment of representative sea anemone toxins that contain a β-defensin motif. Disulfide connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown.

APETx1 is a potent blocker (IC_{50} 34 nM) of the human ether a go-go related gene (hERG) K^+ channel. The toxin induces a shift in the voltage dependence of both activation and inactivation, resulting in a block of potassium currents (107). In addition, it was recently shown (110) that APETx1 can inhibit the conductance of Na_V channels (Na_V1.2–Na_V1.6 and Na_V1.8). Contrary to what is reported for hERG modulation, APETx1 does not change the voltage dependence of activation or steady-state inactivation of Na_V channels.

Na_V type 1 and 2 are the largest Na_V channel toxins from sea anemones, with 46–54 residues. Although they are divided into two types, they share up to 50% sequence identity. Moreover, due to the fact that certain toxins from *Halcurias sp., Nematostella vectensis* and *Condylactis gigantean* resemble both type 1 and 2 sequences (113, 114), Moran and co-workers (89) have suggested that this classification should be revaluated (Figure 1.6). The 3D structures of both type 1 and 2 consist of an anti-parallel β-sheet composed of four β-strands and a highly flexible loop, which has been named the 'Arg14 loop', because Arg14 is the most conserved residue (Figure 1.6) (115-119). Site-directed mutagenesis of AP-B (Δ-actitoxin-Axm1b) revealed that the flexibility of this loop is important for the selectivity and binding of these toxins to Na_V channels.

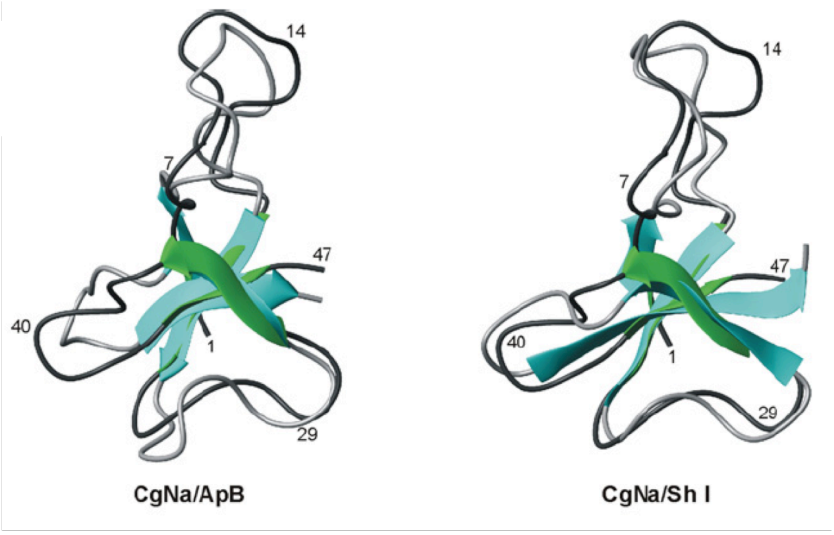


Figure 1.6 – Superimposed cartoon representation of the structures of CgNa (Δ-actitoxin-Cgg1a) and ApB (left) and CgNa and ShI (Δ-stichotoxin-She1a) (right). CgNa is coloured green and dark grey. Figure modified from Salceda *et al.* (119).

Na_V type 1 sea anemone toxins are highly potent modulators of Na_V channels. These toxins bind to a region of the channel named receptor site-3 (i.e., the extracellular S3−S4 loop in domain 4), which is also recognized by scorpion α-toxins (120). Given the close evolutionary relationship between crustaceans (sea anemones prey and predator) and insects, sea anemone toxins also have a profound effect on insect Na_V channels. For this reason, these peptides have been considered as lead compounds in the development of bioinsecticides (91). Anthopleurins (type 1 Na_V channel toxins isolated from the genus *Anthopleura*) and related type 1 Na_V channel toxins have also been considered for therapeutic applications. It was believed that these

toxins could be used in the cardiovascular field but these expectations were not met, in part because of arrhythmogenic activity in the heart (121).

The Na_V type 4 sea anemone toxin family is comprised of only two isotoxins, Calitoxin I (Δ-hormotoxin-Cpt1a) and II (Δ-hormotoxin-Cpt1b), both isolated from *Calliactis parasitica* (122, 123). These toxins contain 46 amino acid residues with only a single amino acid difference (Glu8 versus Lys8). In crustacean nerve muscle preparations, they interact with axonal, but not with muscle, membranes, inducing a massive release of neurotransmitter that causes a strong muscle contraction. They resemble Na_V type 1 and 2 toxins with regard to chain length and the number of disulfide bridges (three) but not in amino acid sequence, sharing only ~45% sequence identity. Evaluation of the Na_V subtype selectivity of these toxins and isolation of more members of this type should help in future classification of this group.

APETx2 from *Anthopleura elegantissima* was the first ASIC-targeting peptide isolated from sea anemone venom and only the second from any venomous animals (106). APETx2 (π-actitoxin-Ael2b) is a 4558 Da peptide (42 residues) that selectively blocks ASIC3 homomeric channels (IC_{50} 63 nM) and the ASIC3-containing heteromers ASIC2b-ASIC3 (IC_{50} 117 nM), ASIC1b-ASIC3 (IC_{50} 0.9 μM) and ASIC1a-ASIC3 (IC_{50} 2 μM). The structure of APETx2 was determined using two-dimensional ^1H NMR spectroscopy using the native toxin (104) (Figure 1.5). It belongs to the disulfide-rich all-β structural family with a fold typical of the β-defensin family (124). To date, three peptides have been isolated from sea anemone venoms that target ASIC channels and, interestingly, they do not have the same structural fold. This structural diversity highlights sea anemone venoms as excellent sources of novel ion channel modulators.

1.3.4.3 Boundless β-hairpin

Osmakov and collegues (125) reported three peptides with uncommon β-hairpin structure isolated from venom of the sea anemone *Urticina grebelnyi*. One of these peptides, Ugr9-1 (π-actitoxin-Ugr1a), reversibly inhibits both transient and sustained currents mediated by human ASIC3 channels. NMR spectroscopy revealed that Ugr9-1 has an unusual structure, stabilized by two disulfide bonds, with three classical β-turns and a twisted β-hairpin devoid of interstrand disulfide bonds (Figure 1.7). Although the authors suggested that this represents a novel peptide fold, which they named the boundless β-hairpin (BBH), other sea anemone toxins with similar disulfide framework had in fact been reported previously (126, 127). These toxins belong to K_V type 4, which is comprised of three toxins, Bcg-III-23.41 and SHTX-1/SHTX-2

(κ-stichotoxin-Shd1a/b) (Figure 1.8). The activity of SHTX-I was indirectly assayed by competitive inhibition of the binding of ^{125}I- α -dendrotoxin to rat synaptosomal membranes but its channel blocking specificity is not yet known. The only difference between SHTX-I and II is a posttranslational modification of Pro6 in SHTX-II to a hydroxyproline in SHTX-I.

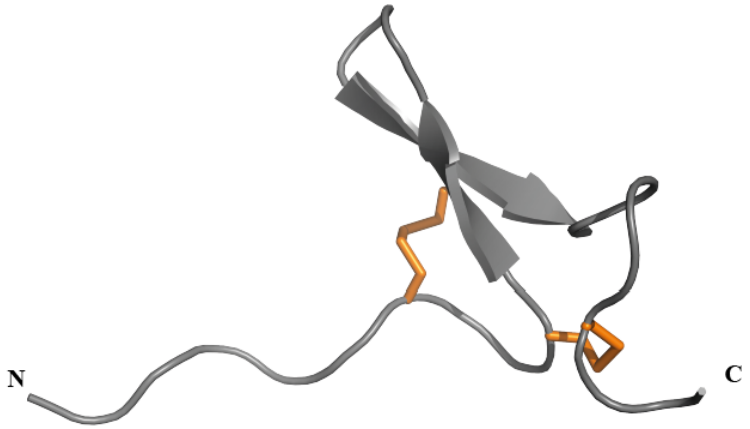


Figure 1.7 – 3D structure of the sea anemone toxin Ugr9-1 (PDB accession code 2LZO). The two disulfide bonds are represented by orange tubes.

Another toxin with similar framework was found in the venom of *Stichodactyla duerdeni*, and named U-SHTX-Sdd1. Although its pharmacological activity remains to be determined, U-SHTX-Sdd1 was the first sea anemone toxin described with an *O*-linked hexose-*N*-acetyl posttranslational modification, in this case of the N-terminal threonine (128). Recently, a novel BBH peptide that produces a significant potentiating effect on allyl isothiocyanate- and diclofenac-induced TRPA1 currents was isolated from venom of the sea anemone *Metridium senile* (129). Ms 9a-1 acts as a positive modulator of TRPA1 *in vitro* but did not cause pain or thermal hyperalgesia when injected into the hind paw of mice. The Ms 9a-1 protein precursor also encodes two homologous toxins named Ms 9a-2 and Ms 9a-3 that are distinguished from Ms 9a-1 by a shorter C-terminal tail and a non-homologous region between the 2nd and 3rd Cys residues (Figure 1.8).

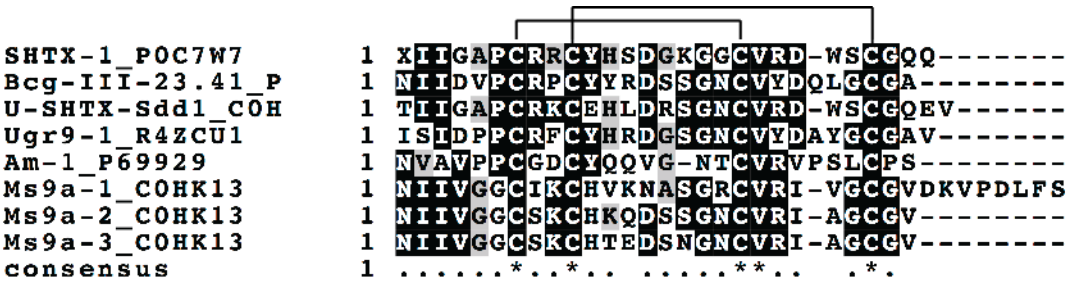


Figure 1.8 – Alignment of representative sea anemone toxins that adopt the BBH motif. Disulfide connectivities are indicated. Amino acid identities (black boxes) and similarities (grey boxes) are shown.

1.3.4.4 EGF-like peptides

Gigantoxin I (ω-stichotoxin-Sgt1a) is a peptide toxin from *Stichodactyla gigantea* that has homology with mammalian epidermal growth factor (EGF). In accordance with this sequence homology, this toxin exhibits EGF activity as evidenced by rounding of human epidermoid carcinoma A431 cells (130). Gigantoxin I also modulates the activity of TRPV1 channels (131), but this activity results from the involvement of the epidermal growth factor receptor/PLA_2/arachidonic acid/lipoxygenase pathway in indirect activation of TRPV1. Gigantoxin I was the first toxin described that induced this effect. Because of this, Cuypers and co-workers (131) suggested that this new group of toxins might help to better understand the regulation of TRPV1 channels.

1.3.4.5 Inhibitor Cystine Knot fold

Although the 3D structure of sea anemone ICK peptides remain to be confirmed, two different types of sea anemone toxins display cysteine patterns characteristic of the ICK structural scaffold, namely K_V type 5 toxins and the ASIC toxin PhcrTx1 (π-phymatoxin-Pcf1a) (Figure 1.9) (88).

PhcrTx1, the first peptide characterized from venom of *Phymanthus crucifer*, is a 32-residue peptide with three disulfide bonds (88). Although its disulfide framework and 3D structure remain to be determined, the distribution of cysteines has the "classic" ICK signature (i.e., $CX_nCX_nCCX_nCX_nC$, where X is any amino acid and *n* indicates a variable number of amino acid residues). If confirmed, this would be the first sea anemone toxin reported to contain an ICK motif. PhcrTx1 reversibly inhibits ASIC currents in rat dorsal root ganglia neurons with an IC_{50} of 100 nM.

K_V type 5 toxins contain only a single recently reported toxin, namely BcsTx3 (κ-actitoxin-Bcs4a) from *B. caissarum* (132), although there are *putative* members of this family from *Nematostella vectensis* (NvePTx1) and *Metridium senile* (MsePTx1). BcsTx3 is a single-chain peptide containing 50 residues crosslinked by four disulfide bridges. Its cysteine framework is similar to an ICK with an additional disulfide bridge, although like PhcrTx1, the structure of this toxin remains to be determined. BcsTx3 has high affinity for *Drosophila* Shaker IR channels (IC_{50} 94 nM) over $K_V1.2$ (IC_{50} 173 nM), $K_V1.3$ (IC_{50} 1007 nM) and $K_V1.6$ (IC_{50} 2246 nM) channels.

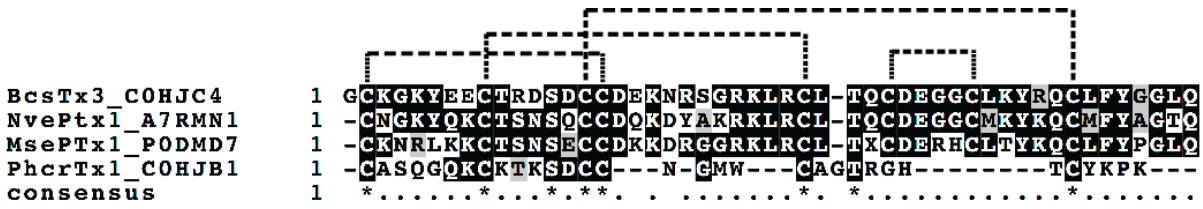


Figure 1.9 – Alignment of sea anemone toxins that likely adopt an ICK motif. Predicted disulfide connectivities are indicated. Amino acid identities (black boxes) and similarities (grey boxes) are shown.

1.3.4.6 ShK motif

The ShK motif was named after the toxin ShK (κ-stichotoxin-She3a), which was identified from the venom of *Stichodactyla helianthus*. This toxin belongs to K_V type 1, which is comprised of toxins with 35–37 amino acid residues that block K_V1 channels. The solution structure of ShK (133) consists of two short α-helices encompassing residues 14–19 and 21–24, and an N-terminus with an extended conformation up to residue 8, followed by a pair of interlocking turns that resemble a 3_{10}-helix (Figure 1.10). It contains no β-strands.

ShK blocks $K_V1.1$, $K_V1.3$ and $K_V1.6$ channels with picomolar potency, and three other channels ($K_V1.2$, $K_V3.2$ and $K_{Ca}3.1$) with nanomolar potency (134). The surface of ShK involved in binding to K_V channels has been probed using alanine scanning mutagenesis and selected toxin analogues (135). These studies revealed that two residues, Lys22 and Tyr23, are crucial for ShK activity, as also found subsequently for BgK toxin (136), while other residues contribute to the K_V channel-binding surfaces. However, Gasparini and colleagues (137) proposed that the Lys-Tyr motif be more broadly defined as a lysine and a neighbouring hydrophobic residue. ShK blocks K_V channels by binding to a shallow vestibule at the outer entrance to the ion conduction pathway and occluding ion entrance to the pore (138-140).

Although it inhibits several K_V subtypes, ShK most potently blocks $K_V1.3$ with an IC_{50} of ~10 pM. $K_V1.3$ plays a critical role in subsets of T and B lymphocytes implicated in autoimmune disorders, and ShK has therefore been studied as potential immunomodulator for therapy of autoimmune diseases. ShK analogs were developed to be more specific to $K_V1.3$ (138). One of these analogs, ShK-186, is being developed as a therapy for autoimmune diseases. ShK-186 (dalazatide) was well tolerated in a recently completed human phase 1A safety trial (141), and it is being advanced by Kineta Inc. (Seattle, WA, USA) into Phase 2 clinical trials (142).

Similar to ShK, BgK inhibits $K_V1.1$, $K_V1.2$, $K_V1.3$, $K_V1.6$ and K_{Ca} currents at nanomolar concentrations (136), although it has a high affinity for $K_V1.1$ (K_d = 6 nM for Kv1.1, 15 nM for $K_V1.2$, 10 nM for Kv1.3). Beraud and co-workers (143) proposed that $K_V1.1$ blockade has broad therapeutic potential in neuroinflammatory diseases (multiple sclerosis, stroke, and trauma). They therefore used an analog of BgK, BgK-F6A, which has the same high affinity for $K_V1.1$ (IC_{50} 0.72 nM) but decreased affinities for $K_V1.2$ (IC_{50} 400 nM) and $K_V1.3$ (IC_{50} 800 nM), to provide preclinical evidence that $K_V1.1$ blockers could be used to treat neuroinflammatory diseases.

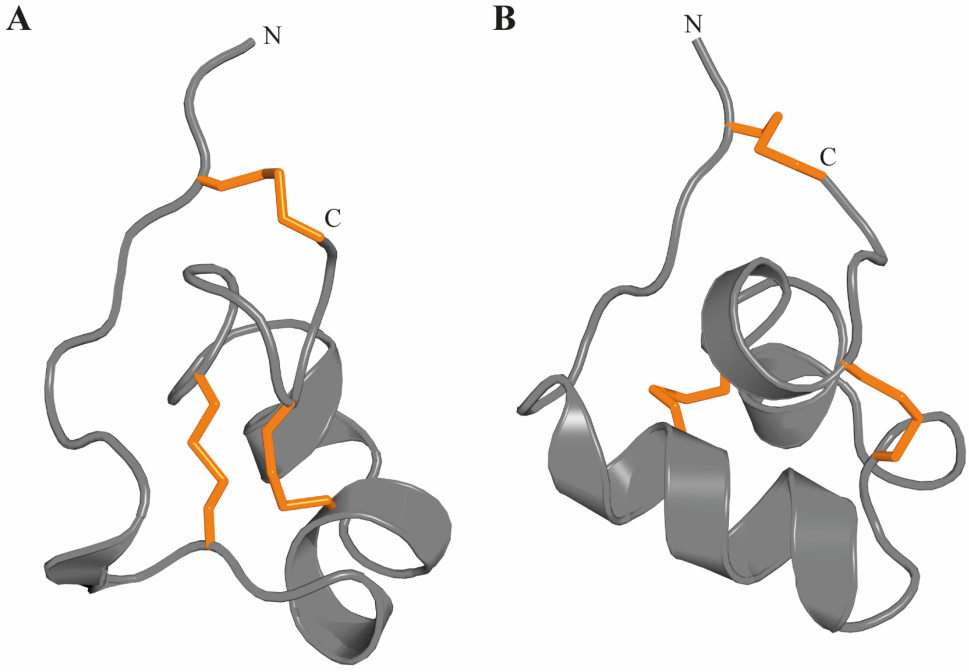


Figure 1.10 – 3D structure of Kv type 1 sea anemone toxins. (A) Structure of BgK (PDB accession code 1BGK). (B) Structure of ShK (PDB accession code 1ROO). The three disulfide bonds are represented by orange tubes.

1.3.4.7 Kunitz-domain

The Kunitz-type protease inhibitors are the best-characterized family of serine protease inhibitors, probably due to their abundance in several organisms. The Kunitz-type motif consists of a peptide of ~60 amino acid residues stabilized by three disulfide bridges (C_1–C_6, C_2–C_4, C_3–C_5). Its 3D structure is characterized by an α/β/α motif (144) (Figure 1.11) with a hydrophobic core. The first reports on the existence of protease inhibitors in sea anemones date from the 1970s (145, 146). To date, protease inhibitor peptides and neurotoxins have been isolated from sea anemone whole bodies, tentacles, secreted mucus and aggressive organs such as acrorhagi, which is present in some species from the family Actiniidae (147). Several protease inhibitors have already been isolated or partially purified and characterized from several sea anemone species.

K_V type 2 sea anemone peptide toxins block K_V1 channel currents, although with much less potency than K_V type 1 toxins (31, 148, 149). Their biological role is still unclear. It is supposed that these protease inhibitors could: (1) defend sea anemones from the proteases of their victims; (2) protect the toxins injected into prey or predators from degradation by host proteases; (3) act on the regulation of digestive mechanisms, including self-digestion by their own enzymes or by those of symbiotic microorganism; (4) also, due to their dual activity, they could also be used to paralyze prey (150). The sea anemone kalicludines (AsKC1 to AsKC3, κπ-actitoxin-Avd3b-d) from *Anemonia sulcata*, APEKTx1 (κπ-actitoxin-Ael3a) from *Anthopleura elegantissima*, SHTX-3 (κπ-stichotoxin-Shd2a) from *Stichodactyla haddoni*, and Sh1 (δ-SHTX-She1a) from *Stichodactlyla helianthus*, are examples of toxins with both protease inhibitor and potassium channel blocking activities (127, 149, 151).

Figure 1.11 – 3D structure of the Kv type 2 sea anemone toxin ShPI-I (π-stichotoxin-She2a) (PDB accession code 3OFW). Peptides of this type are homologous to Kunitz-type inhibitors of serine proteases. The three disulfide bonds are represented by orange tubes.

Many protease inhibitors were isolated from a body extract of *H. crispa*, but only a few have been fully characterized. Protease inhibitors were obtained from a water-ethanol extract of *H. crispa*, and the primary structure was elucidated for one of them, named Kunitz-type trypsin inhibitor IV or Jn-IV (152). Four trypsin inhibitors were subsequently isolated (InI–InIV), one of which (InI) was partially characterized. Later on, also from *Heteractis crispa* extract, a Kunitz-type toxin designated InhVJ was isolated (153, 154). InhVJ is highly specific toward trypsin and α-chymotrypsin and does not inhibit other serine (such as thrombin, kallikrein and plasmin), cysteine (papain) or aspartic (pepsin) proteases. Recently, APHC1 (τ-stichotoxin-Hcr2b) (28) and two homologous peptides (APHC2 and APHC3, τ-stichotoxin-Hcr2b and -Hcr2c) (155) were characterized from *H. crispa*. The APHC toxins contain 56 residues and have high sequence similarity to the Kunitz-type protease inhibitor family (32). Consistent with this, APHC1 and APHC3 are weak inhibitors of serine proteases (156, 157). Its main activity is related to an effect on TRPV1 channels (28). TRPV1 was previously described as a pharmacological target of other cnidarians venoms, such as jellyfish (158). However, APHC1 was the first peptidic TRPV1 modulator isolated from sea anemone venom (156). The primary structure of APHC1 (UniProt B2G331) and APHC3 (UniProt C0HJF4), differ in only four amino acid residues (Figure 1.12). These substitutions result in differences in their ability to modulate TRPV1. 200 nM APHC1 inhibits ~32% of capsaisin-induced currents 200 nM (156), while APHC3 has a lower inhibitory effect (25%) at higher concentrations (300 nM) (159).

Nevertheless, both toxins have antinociceptive and analgesic activity *in vivo* at doses of 0.01–0.1 mg/kg due to their inhibition of TRPV1. Andreev *et al*. (159) suggested that APHC1 and APHC3 might represent a new class of TRPV1 modulators that produce a significant analgesic effect without hyperthermia.

Figure 1.12 – Aligment of APHC1, 2 and 3 (UniProt B2G331, C0HJF3 and C0HJF4, respectively). Cysteines are highlighted in bold, mutations are marked with dots, and conserved positions are marked with *.

1.3.4.8 SCRiPs

SCRiPs were originally identified as genes unique to reef-building corals (Scleractinia) that are downregulated during heat stress (160). Given the similarity in the temporal expression pattern they share with galaxin, a key protein involved in the biomineralization process (161), SCRiPs were implicated in calcification of the coral skeleton (160). However, Jouiaei *et al*. (90) showed that SCRiPs from coral reef (*Acropora millepora*) cause profound neurotoxic effects in fish and it is most likely that they are employed as neurotoxins. Moreover, BLAST searches uncovered SCRiP homologues in the sea anemones *Anemonia viridis* and *Metridium senile* (90). Recently, the first SCRIP was isolated from a sea anemone by Logashina and co-workers (23), who isolated and characterized a peptide from Urticina eques, Ueq 12-1 (τ-AnmTx Ueq 12-1) (Figure 1.13). This study confirmed that SCRiPs act as toxins as predicted by Jouiaei and colleagues; Ueq 12-1 was found to be a bifunctional molecule that exhibits both antimicrobial and TRPA1 potentiating activity, producing an analgesic effect in animal models of pain (23).

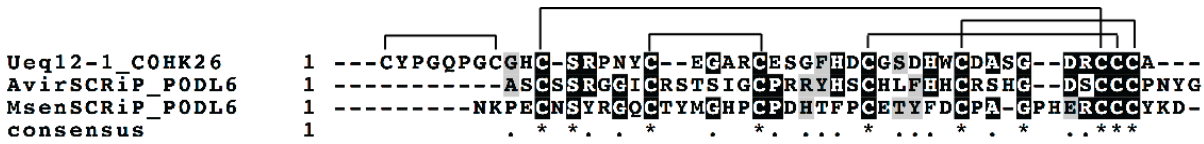


Figure 1.13 – Representative sea anemone SCRiP toxins. Disulfide connectivities are indicated above the sequence alignment. Amino acid identities (black boxes) and similarities (grey boxes) are shown.

SCRiPs contain 8–10 cysteine residues, including a characteristic triplet of cysteines near the C-terminus (Figure 1.13). The 3D structure of Ueq 12-1 reveals that SCRiPs are organized into

a W-shaped structure (Figure 1.14), the core of which is formed by a three-stranded antiparallel β-sheet, a small two-stranded parallel β-sheet, and one turn of a 3_{10} helix stabilized by 4–5 disulfide bridges (C1–C2, C3–C8, C4–C7, C5–C9 and C6–C10). The surface of the peptide is polar without pronounced clusters of positively or negatively charged side chains (23).

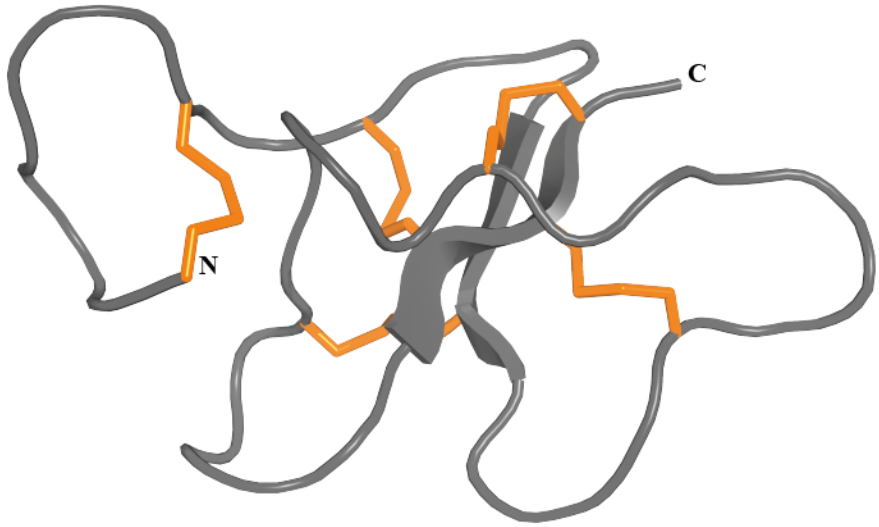


Figure 1.14 – 3D structure of the sea anemone toxin Ueq 12-1 (PDB accession code 5LAH). The three disulfide bonds are shown as orange tubes.

1.4 Summary and significance of project

As detailed in this introduction, sea anemones have complex venoms that they depend on for defense, prey capture, and competitor deterrence. These venoms contain an impressive variety of both proteinaceous and non-proteinaceous substances. Peptide toxins in sea anemone venoms are known to act on a wide range of ion channels, including ASIC, TRP, Na_V and K_V channels.

Even though significant progress has been made over the past few years in understanding the ecology, evolution and composition of sea anemone venom, the venom of only ~40 of the more than 1200 recorded species of sea anemones have been investigated. In addition, like many venomous lineages, characterization of toxic components in sea anemones has been done mostly through an opportunistic approach, focusing on peptides and taxa that are easily accessible. This means that the venoms of these animals contain a huge diversity of compounds that remain to be investigated. Access to the composition of the venom is essential for understanding its ecological role and the evolutionary processes involved in toxin recruitment.

Improvements in the techniques used and investigation of a wider array of species will facilitate the discovery of new compounds, some of which might be useful as pharmacological tools, or as leads for development of drugs and insecticides.

1.5 Aims

The primary aim of this thesis is to provide the first comprehensive insight into the evolution, diversification and biodiscovery potential of sea anemone venoms. I used four distinct approaches to fulfill these aims:

1. The overall composition of Australian sea anemone venoms was mapped using an integrated proteomic/transcriptomic approach, the first time this approach has been used for any sea anemone venom;
2. Bioinformatic analysis of the transcriptomic and proteomic data was undertaken to understand the molecular evolution of sea anemone toxins. In addition, the pharmacological activity and three-dimensional structure of selected novel toxin scaffolds was determined using a combination of electrophysiology and NMR spectroscopy.
3. Toxin function was inferred by examining the tissue distribution of toxins using imaging mass spectrometry approach;

This thesis has contributed to our understanding of the evolution and ecological role of sea anemone venoms, and provided data that will allow better classification of sea anemone toxins and the discovery of new toxin folds.

CHAPTER 2

Revisiting venom of the sea anemone *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus

MANUSCRIPT PUBLISHED IN JOURNAL OF PROTEOMICS, JULY 2017

Abstract

More than a century of research on sea anemone venoms has shown that they contain a diversity of biologically active proteins and peptides. However, recent omics studies have revealed that much of the venom proteome remains unexplored. We used, for the first time, a combination of proteomic and transcriptomic techniques to obtain a holistic overview of the venom arsenal of the well-studied sea anemone *Stichodactyla haddoni*. A purely search-based approach to identify putative toxins in a transcriptome from tentacles regenerating after venom extraction identified 508 unique toxin-like transcripts grouped into 63 families. However, proteomic analysis of venom revealed that 52 of these toxin families are likely false positives. In contrast, the combination of transcriptomic and proteomic data enabled positive identification of 23 families of putative toxins, 12 of which have no homology known proteins or peptides. Our data highlight the importance of using proteomics of milked venom to correctly identify venom proteins/peptides, both known and novel, while minimizing false positive identifications from non-toxin homologues identified in transcriptomes of venom-producing tissues. This work lays the foundation for uncovering the role of individual toxins in sea anemone venom and how they contribute to the envenomation of prey, predators, and competitors.

Biological significance

Proteomic analysis of milked venom combined with analysis of a tentacle transcriptome revealed the full extent of the venom arsenal of the sea anemone *Stichodactyla haddoni*. This combined approach led to the discovery of 12 entirely new families of disulfide-rich peptides and proteins in a genus of anemones that have been studied for over a century.

Highlights

- Proteomic/transcriptomic analysis revealed the venom proteome of *Stichodactyla haddoni*
- We identified 12 new families of disulfide-rich peptides/proteins in *S. haddon*i venom
- There was a poor correlation between toxin transcript levels and abundance in venom
- Proteomics data was essential to reliably identify toxins in tentacle transcriptome

2.1 Introduction

Sea anemones belong to the phylum Cnidaria, the oldest extant lineage of venomous animals, with molecular and fossil data placing their origin prior to the Ediacaran period ~750 million years ago (162). Sea anemones are solitary marine invertebrates that have achieved great ecological success despite their structural simplicity. Essentially laminar organisms, their two-dimensional epithelial construction has shaped both behavioral and physiological responses and led to great diversity, as evidenced by their presence in all marine habitats and at all depths and latitudes.

Like other cnidarians, sea anemones secrete venom from specialized cells known as penetrant nematocysts that contain an eversible hollow tubule filled with venom (162). Contact with prey induces explosive eversion of the tubule, which penetrates the target organism and discharges the venom. Sea anemones use venom for predation, defense, and competitor deterrence (163). Their venom has been studied in more detail than other cnidarians, beginning with pioneering work by the French physiologist Charles Richet who in 1903 partially purified and characterised two active components from tentacular extracts of the sea anemones *Actinia equina* and *Anemonia sulcata* (164, 165). However, despite Richet's work and the development of chromatographic methods that enable separation of venom components, most investigators of marine toxins were still studying crude extracts from whole animals or tentacles up until the 1970s (166). It was not until 1976 that the first amino acid sequence was reported for a sea anemone venom peptide (167). In 2007, the study of sea anemone venom entered the 'omics' era when the genome of *Nematostella vectensis* was reported (85), the first for any venomous animal. Since then, sea anemone venoms have been investigated using both transcriptomics (168) and proteomics (10), resulting in the discovery of new types of toxins (88, 94) and tissue-specific patterns of toxin expression (10, 169, 170). However, an integrated proteomic/transcriptomic approach has not yet been used to characterise the overall composition of any sea anemone venom, hampering both our understanding of the evolution of their venoms and the complexity of their venom arsenal.

Most studies of sea anemone venoms have used large, easily accessible species, and/or species harmful to humans. Many studies employed carpet anemones of the genus *Stichodactyla*, which is comprised of five species (*S. gigantea*, *S. haddoni*, *S. helianthus*, *S. mertensii*, and *S. tapetum*). Although these species are restricted to the tropics, *Stichodactyla* are common in the aquarium trade, which, along with their large size and venom yield, makes them attractive subjects for venom-based studies. However, despite >100 publications on *Stichodactyla*

venom, sequences have been reported for only 17 peptide/protein toxins from four species (127), a figure that is at odds with the abundance and diversity of toxins reported for most marine and terrestrial invertebrate venoms (171-175).

Sea anemone venoms contain a variety of both proteinaceous (peptides and proteins) and non-proteinaceous substances (e.g., purines, quaternary ammonium compounds, and biogenic amines) (176). The proteinaceous toxins that have been well characterized can be organized into three major groups: phospholipase A_2, cytolysins, and neurotoxins (177). Both cytolysins and neurotoxins are functional terms, and these two groups actually comprise multiple protein and peptide families — the cytolysin group currently includes five protein families ranging in size from 5 to 80 kDa, while the neurotoxins include cysteine-rich peptides distributed across eight unique structural scaffolds. However, recent proteomic (178) and transcriptomic (177) studies have revealed a large diversity of novel, uncharacterised compounds in sea anemone venoms.

Here, we highlight the power of using a combined transcriptomics/proteomics approach to provide a holistic overview of the complexity of the venom arsenal of sea anemones. Application of this approach to a representative of the well-studied genus *Stichodactyla* uncovered twelve new families of putative toxins, illustrating how much there is still to be learnt about sea anemone venoms.

2.2 Materials and methods

2.2.1 Specimen and Venom Collection

Sea anemones were collected (based on species identification by marine biologists B.M. and E.A.B.U.) at North Stradbroke Island, Queensland, Australia (27°15 S, 153°15 E) then housed in aquaria at the University of Queensland. Venom was obtained after a starvation period of at least 48 h. Briefly, the sea anemone was rinsed, placed in a minimal volume of artificial seawater, and the nematocysts induced to discharge by electrical stimulation (179). The water, which contains the venom, was lyophilized and then the venom was desalted by dialysis at 4°C (Biotech Cellulose Ester membrane, 0.1–0.5 kDa cut-off; Spectrum Labs, USA).

2.2.2 2DE Analysis

Desalted venom (0.4 mg) was solubilized in 125 µL of DeStreak Rehydration Solution (GE

Life Sciences, USA). The sample was mixed and centrifuged to pellet insoluble material, then 1% Immobilized pH Gradient (IPG) buffer (pH 3–10 NL; GE Life Sciences) and 10 mM DTT were added to the supernatant before loading onto isoelectric focusing (IEF) strips (ReadyStrip, non-linear pH 3–10, 7 cm; Bio-Rad, USA) for 24 h passive rehydration. Proteins were electrophoresed in an Ettan IPGphor3 IEF system (GE Life Sciences, USA) under the following conditions: 100 V for 1 h, 300 V for 200 Volt-hours (V-h), 300 to 1000 V for 300 V-h, 1000 to 5000 V for 4000 V-h, and 5000 V for 1250 V-h. The IPG strip was then equilibrated for 10 min in reducing equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 1.5% DTT) followed by a second incubation for 20 min in alkylating equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 2% iodoacetamide). The IPG strip was then embedded on top of a 12.5% polyacrylamide gel (PowerPac Electrophoresis unit; Bio-Rad) and covered with 0.5% agarose. Second dimension electrophoresis was performed at 4°C for 1 h at 150 V per gel. The resulting gel was stained overnight with 0.2% colloidal Coomassie brilliant blue G250 (34% methanol, 3% phosphoric acid, 170 g/L ammonium sulfate, 1 g/L Coomassie blue G250), then destained in 1% acetic acid/H_2O. Visible spots were subsequently picked from the gel and digested overnight at 37°C using sequencing-grade trypsin (Sigma, USA). Briefly, gel spots were washed with ultrapure water, destained (40 mM NH_4CO_3/50% acetonitrile (ACN)) and dehydrated (100% ACN). Gel spots were rehydrated in 10 μL of 20 μg/ml proteomics-grade trypsin (Sigma-Aldrich) and incubated overnight at 37°C. Digests were eluted by washing the gel spots for 30 min with each of the following solutions: 50 μL 50% ACN/1% formic acid (FA), followed by 50 μL 70% ACN/1% FA. The samples were then dried by evaporation using a vacuum centrifuge and reconstituted in 20 μL of 1% FA prior to analysis by LC-MS/MS.

2.2.3 HPLC

Reversed-phase (RP) HPLC analysis of crude *S. haddoni* venom was performed using a Shimadzu Prominence system. Venom (1 mg) was fractionated on a C_{18} column (4.6 × 250 mm, 5 μm particle size, 300 Å pore size) using a flow rate of 1 ml/min and the following gradient of solvent B (0.043% trifluoroacetic acid (TFA) in 90% ACN) in solvent A (0.043% TFA in water): 10% solvent B for 15 min, 10–65% solvent B over 50 min, 45–70% solvent B over 5 min. Fractions were collected manually, dried by evaporation using a vacuum centrifuge, then prepared for LC-MS/MS analysis as described below.

2.2.4 Protein identification using LC-MS/MS

To identify proteins present in the milked venom we used a bottom-up proteomics approach to analyze the digested 2DE gel spots, RP-HPLC fractions, as well as crude desalted venom. Reduction and alkylation of cysteine residues in venom proteins and peptides was performed as reported previously (180). Reduced/alkylated venom was incubated overnight at 37°C in 10 µL of 40 ng/µL proteomics-grade trypsin (Sigma) in 40 mM NH_4CO_3, pH 8. The digested reduced/alkylated samples were then resuspended in a final concentration of 1% FA and centrifuged for 15 min at 12,000 g prior to LC-MS/MS. For analysis of RP-HPLC fractions and in-gel digests, tryptic peptides were fractionated on an Agilent Zorbax stable-bond C_{18} column (2.1 mm × 100 mm, 1.8 µm particle size, 300 Å pore size) using a flow rate of 180 µl/min and a gradient of 1–40% solvent B (90% ACN, 0.1% FA) in 0.1% FA over 15 min on a Shimadzu Nexera UHPLC coupled with an AB SCIEX 5600 mass spectrometer equipped with a Turbo V ion source heated to 500°C. MS/MS spectra were acquired at a rate of 20 scans/s, with accumulation time of 0.25 ms, resulting in a cycle time of 2.3 s, and optimized for high resolution. Precursor ions with m/z of 300–1,800 m/z, a charge of +2 to +5, and an intensity of at least 120 counts/s were selected, with a unit mass precursor ion inclusion window of ± 0.7 Da, and excluding isotopes within ±2 Da for MS/MS. The crude venom digest was analyzed as above except using a gradient of 1–40% solvent B (90% ACN, 0.1% FA) in 0.1% FA over 60 min.

Mass spectra were searched against predicted coding sequences (CDSs) from the assembled transcriptome (see below) using ProteinPilot v4.5 (AB SCIEX). Searches were run as thorough identification searches, specifying tryptic digestion and the alkylation reagent as appropriate. Biological modifications and amino acid substitutions were allowed in order to maximize the identification of protein sequences from the transcriptome despite the inherent variability of toxins, potential isoform mismatch with the transcriptomic data, and to account for experimental artifacts leading to chemical modifications. We used a stringent detected protein threshold score of 1% false discovery rate (FDR) as calculated by decoy searches. All mass spectrometry and ProteinPilot data were submitted to the ProteomeXchange Consortium via the PRIDE (181, 182) partner repository with the dataset identifier PXD006253.

2.2.5 Transcriptome sequencing, assembly, and expression analysis

Little is known about the regeneration of sea anemone toxins. In order to maximize the toxin-gene expression levels for transcriptome sequencing, we performed a preliminary study on toxin expression levels at 0, 24, 48, and 72 h after venom extraction. Using primers based on available toxin sequences from *Stichodactyla,* PCR experiments indicated that toxin transcripts are robustly expressed at 72 h after venom extraction, consistent with active replenishment of the toxin arsenal. Although the dynamics of toxin regeneration in sea anemones warrants more thorough investigation, such as examining a representative selection of toxin families and distinguishing toxins produced in glandular cells *versus* nematocytes, we decided to sequence the transcriptome of tentacles collected from *S. haddoni* at 72 h after venom extraction. Tentacle tissue from the same animals used for proteomic analyses were collected with tweezers and flash frozen before total RNA was extracted using TRIzol (Life Technologies) and enriched for mRNA using a DynaBeads Direct mRNA kit (Life Technologies). mRNA (350 ng) was supplied to the Institute for Molecular Bioscience Sequencing Facility for library preparation and sequencing. A paired-end cDNA library (180 bp insert size) was prepared using the TruSeq-3 library kit and sequenced on an Illumina NextSeq 500 (mid-output, 150 bp paired-end reads). The resulting reads were trimmed using Trimmomatic (183) to remove adapter sequences and low-quality reads. Window function-based quality trimming was performed using a windowsize of 75 and a windowquality of 34, and sequences with a resulting length of <100 bp were removed. After quality control, paired-end sequences were *de novo* assembled into contigs using Trinity v2.0.6 (184) using default parameters. The relative abundance of each transcript was estimated by mapping the paired trimmed reads back to the transcriptome assembly and calculating values of Transcripts Per Million (TPM) using RSEM (version 1.1.17) (185) compiled for Galaxy (186). Mapped reads were visualized using the Integrated Genome Viewer (187, 188) and polymorphic sites identified if supported by more than five reads or frequency greater than 5%. Raw sequence reads (SRA: SRR5397293) and Trinity-assembled contigs have been deposited with links to BioProject accession number SAMN06670449 in the NCBI BioProject database (www.ncbi.nlm.nih.gov/bioproject/).

2.2.6 Functional annotation of transcriptome

In order to identify potential toxin-like transcripts, we compared the translated transcriptome to all curated animal toxin sequences. Coding sequences (CDSs) were identified using the

Galaxy tool ‘Get open reading frames (ORFs) or coding sequences (CDSs)’ (189). A minimum CDS length cut-off of 30 residues was used to minimize the probability of not identifying short toxin CDSs. BLASTp searches of the resulting translated CDSs against the UniProt animal venom database (190) (accessed on April 2016) were performed with the upper-limit for the E-value set to 1E-3. Candidate toxin-like transcripts were further processed by removing redundant protein sequences using CD-HIT (191) as well as sequences not containing a signal peptide that could be detected using SignalP (version 4.1) (192). Finally, the toxin candidates were classified into categories according to predicted structure and/or function.

2.2.7 Functional annotation of proteomics data

CDSs from the venom transcriptome were used as a protein database for proteomic analyses. LC-MS/MS spectra were searched against the CDS database using ProteinPilot v4.5 (AB SCIEX). Sequences with less than two peptides with 95% confidence were excluded; the selected sequences from ProteinPilot were extracted using Galaxy tools. These sequences were then BLAST searched against the UniProt animal toxin database (www.uniprot.org/program/toxins) and annotated according to the methodology described for the transcriptome annotation. Sequences without matches to the UniProt toxin database were functionally annotated using a combination of BLAST searches against the NCBI non-redundant protein database and InterProScan (193) using Blast2GO (194). Sequences without known functional or structural motifs were considered "unknowns". To identify the full molecular diversity contained within each of the putative toxin families, sequences identified in the milked venom were then BLAST searched against the translated transcriptome. Redundant protein sequences were removed using CD-HIT and only sequences containing a predicted signal peptide were considered as putative toxin candidates.

2.3 Results

2.3.1 Transcriptome sequencing, assembly and annotation

We used next-generation sequencing of poly-A enriched RNA extracted from tentacles that were actively regenerating their venom (see ‘Experimental procedures’) to investigate the diversity of toxin-like sequences in sea anemone venom. *De novo* assembly with Trinity yielded 269,628 contigs, a number comparable to previous Illumina-sequenced sea anemone

transcriptomes assembled using Trinity (20, 170), which translated to over 1.8 million potential CDS. As previously reported for sea anemones (170) and other cnidarians (195), we found that the majority of transcripts were not related to the envenomation process, consistent with the tentacles being composed of much more than just the toxin producing nematocytes and gland cells. Nevertheless, by combining BLAST searches with searches for structural and functional domains, we identified 508 unique sequences as putative toxins of which 68% were proteins and 32% were peptides (i.e., < 75 residues). We were able to group these toxins into 63 families (Figure 2.1).

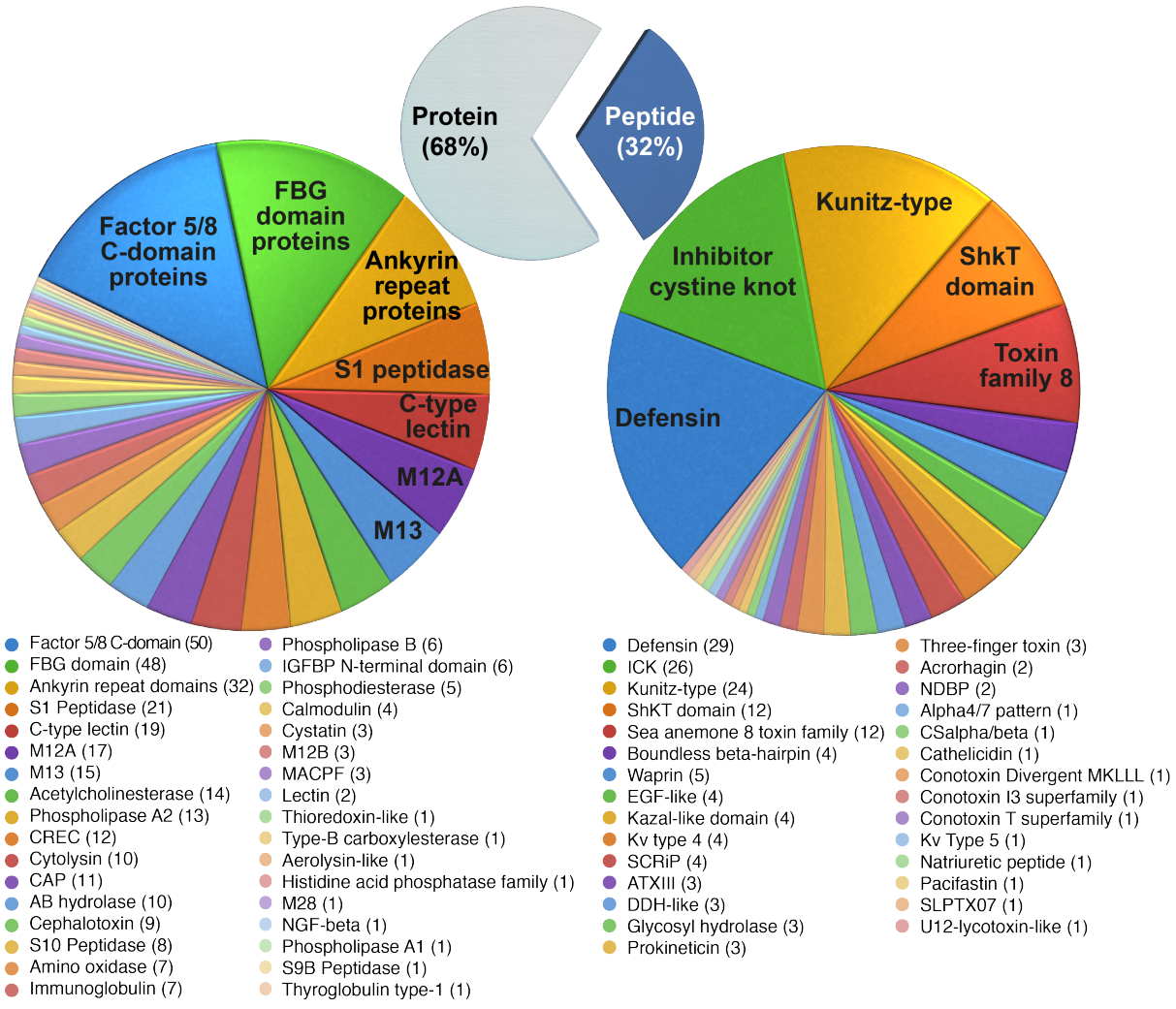


Figure 2.1 – Putative toxin families identified in the transcriptome of *S. haddoni* by BLAST search against UniProt. The 508 unique protein sequences with significant BLAST hits to the manually curated list of animal toxins in UniProt (www.uniprot.org/program/Toxins) were sorted into toxin families according to their cysteine scaffold and amino acid sequence. The pie chart shows the proportional contribution of each toxin family to the predicted venom proteome. The number of homologues identified for each protein and peptide family is shown in parentheses after the family name.

Protein toxins: The putative *protein* toxins with the greatest number of homologues identified by BLAST searches were factor V-like (46 homologues), ficolin-like (46 homologues), and latrotoxin-like (32 homologues). However, great caution needs to be exercised in interpreting such hits. Latrotoxins are the largest and most potent neurotoxins isolated from arthropod venoms. They have a molecular mass of 110–140 kDa (196) and exert a neurotoxic effect on insects or vertebrates by inducing massive neurotransmitter release at presynaptic nerve terminals (197). Despite claims to the contrary, they are found *only* in the venom of widow spiders (i.e., the genera *Latrodectus*, *Steatoda*, and *Parasteatoda*) (198). However, because latrotoxins contain 20 or more ankyrin repeat domains (196), one of the most ubiquitous protein-protein interaction motifs known, there are numerous instances where venom-gland transcripts from a wide variety of venomous animals (199-201) have yielded hits to one or more of these domains, but not other regions of the latrotoxin protein; hence, these hits are not true latrotoxin homologues. This was also the case in the current study — the family of sea anemone proteins with hits to latrotoxin are smaller and highly variable in size (8.1–60.1 kDa) and the sequence homology is confined to the ankyrin repeat domains. Moreover, proteomic analysis of milked venom (see below) did not provide evidence for expression of these proteins in the venom, suggesting that they are not venom toxins. Since these sea anemone proteins contain 1–6 ankyrin repeat domains (InterPro IPR002110; SMART SM00248; Pfam PF00023), we refer to them as "ankyrin repeat proteins" rather than latrotoxin homologues.

Ficolins are large oligomeric lectins (~35 kDa) comprised of an N-terminal domain followed by a long collagen-like stretch that precedes a C-terminal fibrinogen-like globular (FBG) domain (InterPro IPR002181; SMART SM00186; Pfam PF00147). Human ficolins serve as lectin-type pattern recognition receptors that recognise carbohydrates on the surface of microbial pathogens and activate the lectin pathway of the complement system (202). They are rare in animal venoms, with only five reported to date from reptiles (203, 204). In contrast to these reptilian ficolins, the sea anemone proteins with hits to ficolin are highly variable in size (9.2–162.5 kDa), the sequence homology is confined to the FBG domain of ficolin, and proteomic analysis of milked venom did not provide evidence for their expression in the venom. We conclude that they are not venom toxins and refer to them as "FBG domain proteins" rather than ficolin homologues.

Factor V is a key non-enzymatic component of the mammalian coagulation system (205). It is a large protein with domain architecture A1-A2-B-A3-C1-C2 (206). Factor V-like toxins with

the same domain architecture are found in snake venoms, where they interfere with the coagulation and hemostatic systems of envenomated mammalian prey (206). In contrast, the sea anemone transcripts with hits to Factor V are highly variable in size (13.6–223 kDa), the sequence homology is confined to the C-terminal repeat domain found in both Factor V and Factor VIII (Coagulation factor 5/8 C-terminal domain: InterPro IPR000421; SMART SM00231; Pfam PF00754) and proteomic analysis of milked venom did not provide evidence for their expression in the venom. We conclude that they are not venom toxins and refer to them as "Factor 5/8 C-domain proteins" rather than Factor V homologues.

Peptide toxins: The putative *peptide* toxins with the largest number of unique homologues were β-defensins (32 homologues), inhibitor cystine knot (ICK) peptides (27 homologues), and Kunitz-type peptides (24 homologues). β-defensins are antimicrobial peptides that are secreted as part of the innate immune response in a wide range of taxa (102, 207). However, in sea anemone venoms, β-defensin-like peptides have become weaponized to serve as neurotoxins that modify the activity of voltage- and ligand-gated ion channels; this family of peptides includes both K_V type 4 and Na_V type 1, 2 and 4 toxins (103, 105, 208).

The ICK fold is perhaps the most abundant disulfide-rich peptide fold known (209). It is found in a wide diversity of organisms including viruses, plants, fungi and animals where they appear to primarily provide defence against pathogens (209). Due to the stability (210, 211) and evolutionary plasticity (209, 212) provided by the ICK fold, it is also one of the most frequently "weaponized" peptide folds. Hence, ICK-type toxins are found in a wide range of animal venoms, such as those of cone snails (213) and spiders (179), where they act as neurotoxins that affect ion channel function. Only one sea anemone toxin with a putative ICK motif has been purified and characterized (88), although its ICK architecture has not yet been confirmed by structural or chemical analysis.

Kunitz-type peptides have been described from the venom and venom glands of a taxonomically diverse array of venomous animals, where they act as protease inhibitors and ion channel modulators (214, 215). In sea anemones, both TRPV and type II K_V channel toxins assume a Kunitz fold (27). Interestingly, some sea anemone type II K_V channel toxins have dual activity, acting both as potassium channel blockers and inhibitors of serine proteases (163, 216), which enables them to serve both as defense molecules and as neurotoxins that aid in prey immobilization (217).

Although our expression level analysis lacked replicates, it is worth noting that the most abundantly expressed of the putative toxins identified by BLAST were neurotoxin-like peptides and FBG domain proteins (Table 2.1). In contrast, the Factor 5/8 C-domain proteins and ankyrin repeat proteins were not highly expressed. In fact, we found little correlation between the number of homologues within each protein family and its expression level. This correlation was better for peptides, where 4 of the 6 most diverse peptide scaffolds were represented among the top 10 most highly expressed transcripts.

Table 2.1 – Highly expressed toxin transcripts. Ten most highly expressed transcripts encoding putative toxins in the tentacle transcriptome of *S. haddoni* as identified by BLAST search against the manually curated list of animal toxins in UniProt (www.uniprot.org/program/Toxins). Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit.

Transcript ID	*TPM*	*UniProt hit*	*Family*	*E-value*
TR139525_c0_g5_i1_CDS4	11699	P0DMX5	β-defensin	8E-20
TR11907_c0_g1_i1_CDS3	9165	P0DMZ3	Sea anemone 8 toxin family	7E-36
TR121578_c2_g1_i1_CDS2	6421	C0HJF3	Kunitz-type	3E-16
TR131122_c0_g1_i1_CDS5	5757	B2DCR8	Cephalotoxin-like	6E-05
TR27817_c0_g1_i1_CDS7	5671	D8VNS8	Ficolin lectin	2E-06
TR131356_c0_g1_i1_CDS3	4341	P0DMZ6	Sea anemone 8 toxin family	1E-44
TR129689_c0_g1_i2_CDS3	4048	B2G331	Kunitz-type	1E-42
TR90837_c2_g2_i2_CDS3	3681	P0DMX5	β-defensin	3E-20
TR122517_c0_g1_i1_CDS3	3180	D8VNS9	Ficolin lectin	4E-05
TR138219_c4_g1_i1_CDS10	2518	R4ZCU1	Boundless β-hairpin	2E-10

2.3.2 Proteomics of milked venom

To identify which proteins and peptides are present in *S. haddoni* milked venom, we used a combination of 2DE and shotgun analyses of crude and RP-HPLC fractionated venom. In order

to estimate the completeness of our proteomics data, we performed a search for toxins previously reported to be present in *S. haddoni* (127) (see Appendix I). Our RP-HPLC fractionation (Figure 2.2A) and 2DE gel (Figure 2.2B) revealed a complex set of both low and high molecular weight proteins. However, we found little overlap between the putative toxins identified using a purely bioinformatic approach and components identified by the combined proteomic/transcriptomic approach. A ProteinPilot search of all the mass spectrometry data against our translated transcriptome yielded a total of 131 unique coding sequences (see Supplementary Information file S4), of which only 27 were identified as putative toxins during functional annotation of the transcriptome. Moreover, the milked venom of *S. haddoni* contained a high proportion of completely novel toxin families. Of the 131 identified sequences, only 33 showed significant homology to entries in the UniProt toxin database (see Appendix II).

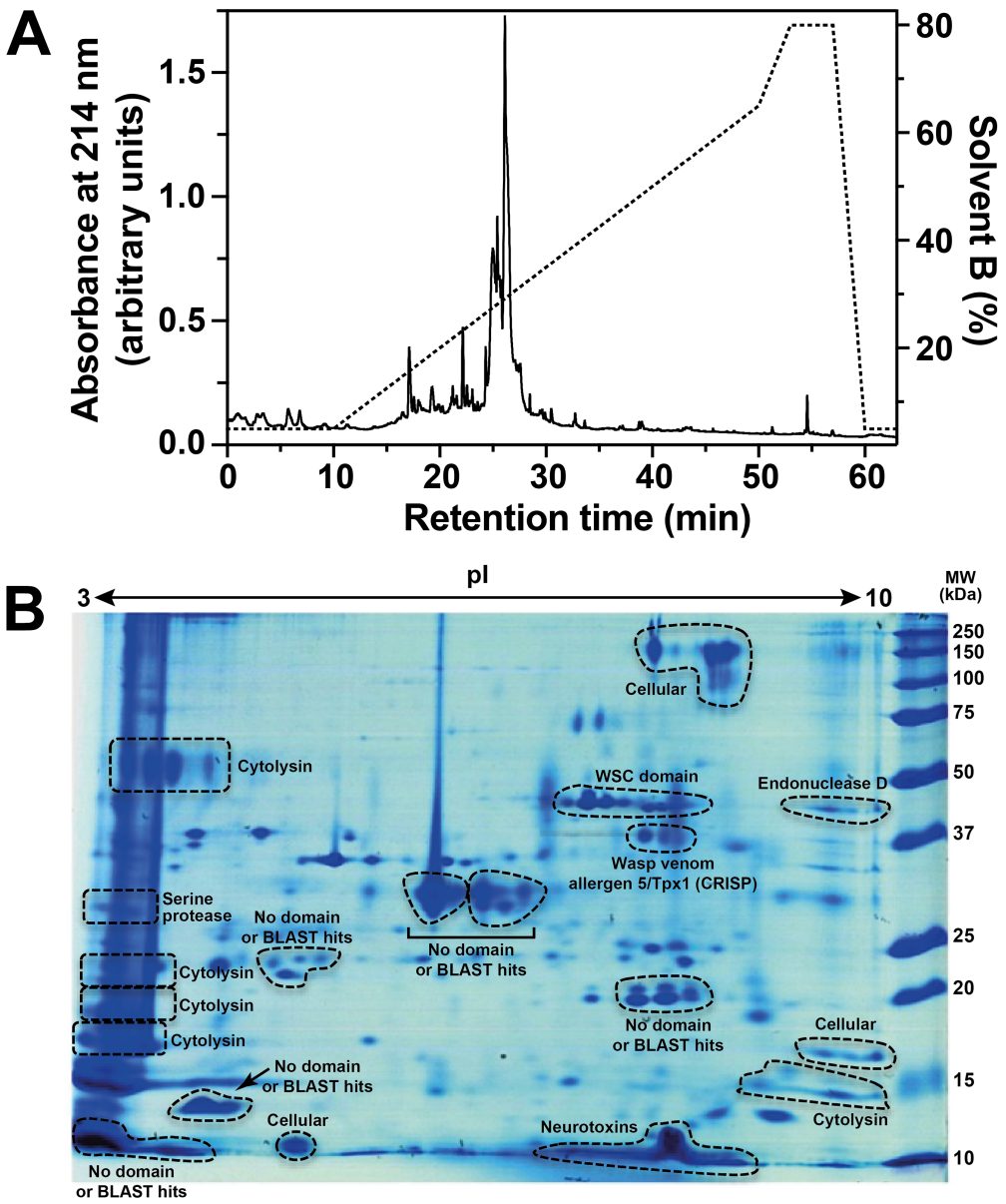


Figure 2.2 – Composition of *S. haddoni* venom. (A) C_{18} RP-HPLC chromatogram of desalted crude *S. haddoni* venom. **(B)** 2DE gel of *S. haddoni* venom; the first dimension was isoelectric focussing (pH 3–10) followed by 12.5% SDS-PAGE. Molecular masses of standards are indicated on right of gel. Proteins identified by in-gel digestion and LC-MS/MS are annotated according to their respective family. Spots with less than two peptides identified in the LC-MS/MS and confidence values below 95% were not annotated.

The final set of potential toxins identified in the milked venom included eleven known toxin families and twelve protein families that could not be assigned a putative function or known family. The putative toxins with the greatest number of homologues identified were β-defensins (7), Kunitz-type peptides (7), and the newly identified toxin family U_{11}-Std (7) (Table 2.2). The identified β-defensins included Na_V type 1, 2 and 4 toxins as well as K_V type 3 toxins, while the Kunitz-type peptides include K_V type 2 and TRPV toxins. Supplementary

Table S1 (not shown in this thesis) provides a complete list of identified putative toxins annotated according to both structural scaffold and traditional sea anemone toxin nomenclature.

Table 2.2 – Novel toxin scaffolds identified in *S. haddoni* venom. Novel protein and peptide families identified in the venom of S. *haddoni* organized into families according to cysteine scaffold and amino acid sequence similarity. 'X' indicates any amino acid.

Toxin family	No. of cysteines	Scaffold	Length of mature peptide	Number of homologues
U_1-Std	4	$x_7Cx_3Cx_5Cx_8Cx_2$	28	4
U_2-Std	6	$xCx_7Cx_6CCx_{12}Cx_{12}Cx_3$	47	2
U_3-Std	6	$x_2Cx_{35}Cx_3Cx_{25}CxCx_{22}Cx$	93	4
U_4-Std	7	$x_5Cx_{47}CCx_{10}Cx_{26}CCx_{83}C$	178	1
U_5-Std	8	$x_3Cx_{13}Cx_3Cx_{13}Cx_5CCxCx_6Cx_4$	56	2
U_6-Std	8	$x_{19}Cx_3Cx_6Cx_3Cx_9Cx_3Cx_6Cx_{21}Cx_{23}$	101	1
U_7-Std	9	$x_9Cx_3CxCx_7Cx_3Cx_6Cx_{19}CxCx_4Cx$	63	1
U_8-Std	10	$x_3Cx_{18}Cx_3CxCx_7Cx_3Cx_6Cx_{19}CxCx_4Cx$	76	1
U_9-Std	10	$Cx_{17}Cx_8Cx_{31}Cx_{35}Cx_7Cx_{14}Cx_8Cx_9Cx_{30}Cx_5$	174	3
U_{10}-Std	12	$x_7Cx_7Cx_{22}Cx_8Cx_{10}Cx_{24}Cx_{41}Cx_7Cx_{14}Cx_8Cx_5Cx_{29}Cx_4$	198	1
U_{11}-Std	18	$x_3Cx_{25}Cx_{19}Cx_{18}Cx_{37}Cx_{12}Cx_{16}Cx_{17}CxCCx_{60}Cx_{11}Cx_3Cx_{46}Cx_6Cx_4Cx_{13}Cx_3$	311	28
U_{12}-Std	20	$x_3Cx_3Cx_6Cx_3Cx_5CCx_{13}Cx_5Cx_4CCx_4Cx_3Cx_6Cx_2Cx_6CCx_9Cx_5Cx_4CCx_2$	103	1

Consistent with the toxin profile suggested by expression analyses of the putative toxins identified by BLAST search against the UniProt toxin database, our proteomic results show

that the primary components in the venom are neurotoxins and one type 2 cytolysin. Furthermore, the main fraction in the RP-HPLC chromatogram (elution time ~26 min) is composed of an EGF-like neurotoxin and a type 2 cytolysin. Among the toxins identified in the venom, we found a Kunitz-type toxin with homology to τ-stichotoxin-Hcr2b (UniProt B2G331) from *Heteractis crispa* to have the highest level of expression (Table 2.3). The second most highly expressed transcript whose translated product was identified in the milked venom encodes a boundless β-hairpin (BBH). Interestingly, the BBH encoded by this transcript is similar to each of the four toxins encoded by the UG precursor (UniProt R4ZCU1) from the painted anemone *Urticina grebelnyi*; however, whereas the UG transcript encodes four homologous toxins that are post-translationally liberated (24), the *S. haddoni* transcript encodes only a single toxin. Curiously, this is the only BBH-encoding transcript whose product was identified in the venom, and one of only two in the entire transcriptome assembly.

As Trinity-based assembly can mask the full diversity of toxin genes (177, 218), we examined the mapped reads to each contig where the ProteinPilot data indicated with high confidence that an amino acid substitution was present in the CDS. Of 11 contigs containing potential amino acid substitutions, six contained hidden polymorphisms (frequency threshold 0.05), four of which result in non-synonymous substitutions in the mature peptide region (Appendix III). Interestingly, of the five additional toxin isoforms identified by re-examining mapped reads, only one matched an amino acid substitution suggested by our proteomic data (TR135650_c2_g2_i1_CDS2; similar to τ-stichotoxin-Hcr2b). It is also interesting to note that while the BBH-encoding contig corresponding to the toxin identified in the venom contained an additional sequence with two polymorphic nucleotide sites, their amino acid sequences were identical.

Table 2.3 – Most highly expressed transcripts identified in milked venom. Ten most highly expressed toxin candidates identified from proteomic analysis of *S. haddoni* venom. Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit.

Transcript ID	*TPM*	*UniProt hit*	*Family*	*E-value*
TR129689_c0_g1_i2_CDS3	4048	B2G331	Kunitz-type	1E-42
TR138219_c4_g1_i1_CDS10	2518	R4ZCU1	Boundless β-hairpin	2E-10

TR130412_c0_g1_i1_CDS3	1777	Q76CA0	β-defensin	2E-56
TR135650_c2_g5_i1_CDS2	1188	B2G331	Kunitz-type	4E-43
TR135650_c2_g2_i1_CDS2	1166	B2G331	Kunitz-type	9E-43
TR135650_c2_g6_i1_CDS2	508	B2G331	Kunitz-type	1E-42
TR104378_c2_g2_i2_CDS3	501	Q9U6X1	Cytolysin type 2	4E-139
TR69080_c0_g1_i1_CDS11	437	B1B5J0	EGF-like	3E-59
TR135650_c2_g8_i1_CDS2	370	B2G331	Kunitz-type	8E-36
TR72284_c0_g1_i1_CDS4	307	A7RMN1	Sea anemone K_V type 5	1E-19

To investigate the full diversity of each toxin family identified in milked *S. haddoni* venom, we conducted a BLAST search with each identified coding sequence against the transcriptome. This revealed that the known toxin families with the greatest number of unique homologues identified in the venom were Kunitz-type (16 homologues), β-defensins (11 homologues) and S1 peptidase (16 homologues) (Figure 2.3). The majority of transcripts in each of these families were, however, not identified in the milked venom, perhaps indicating that some homologues are not venom toxins. Identification of related house-keeping proteins is also a likely scenario for CAP proteins and S1 peptidases, where only one and two out of 14 and 16 homologues, respectively, were identified in the venom.

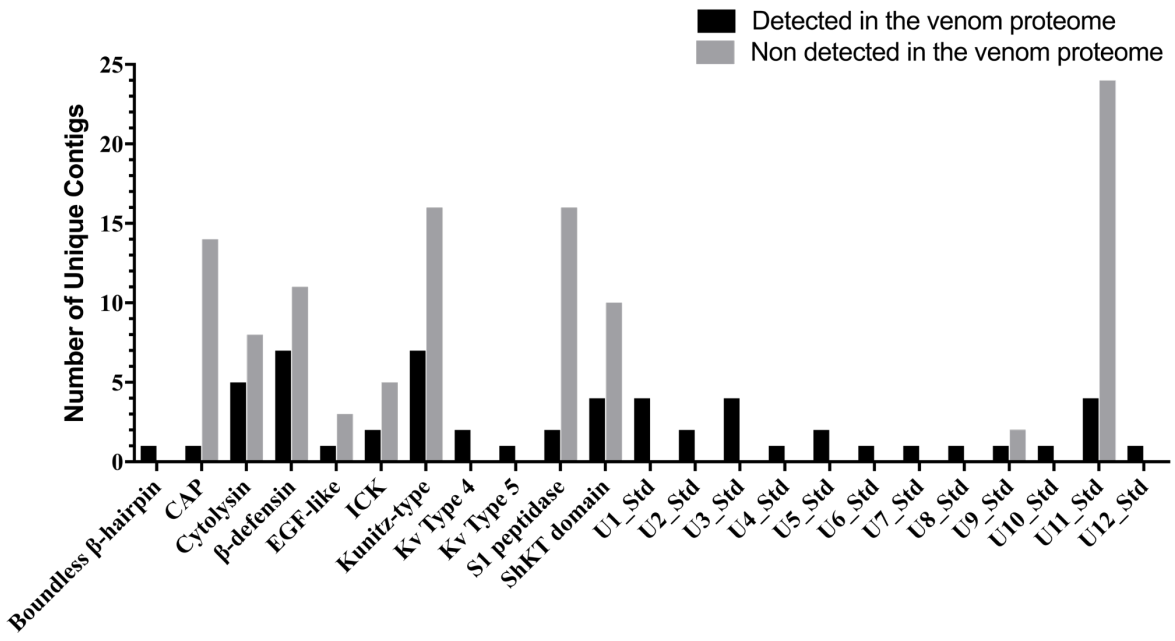


Figure 2.3 – Diversity of putative toxins detected in, or absent from, the venom of *S. haddoni*. Histogram of unique protein sequences within each toxin family identified in the transcriptome that were (black) or were not (grey) identified in proteomic analyses of milked venom. Putative toxin families U_1-Std to U_{12}-Std are new scaffolds as described in Table 2.2.

In addition to the hits obtained from the UniProt toxin database, 12 protein families could not be assigned a putative function or known family. They were named with a U prefix (U = unknown) followed by a subscripted number to indicate the family, then a genus/species identifier (Std) to indicate that they the toxins are from *S. haddoni* (33). All of these novel peptide and protein families are cysteine-rich, and the number of cysteines, cystine scaffold, and predicted length of the mature peptide are summarized in Table 2.2 (Appendix IV). Interestingly, toxins U_1–U_8 were exclusively found in the venom, with no additional non-venom homologues recovered by retrospective BLAST searches of the transcriptomic data (Figure 2.3). These families also included all the novel peptide scaffolds with predicted molecular weight below 10 kDa, suggesting that these are indeed likely novel toxins and not housekeeping peptides.

Although many of the toxin families identified in the venom had highly expressed representatives, we did not find that the putative toxins identified in the venom consistently had higher expression levels than their homologues that were not identified in the venom (Figure 2.4). Moreover, the five most abundantly expressed putative toxins (i.e., β-defensin, sea anemone 8 toxin family, Kunitz-type, cephalotoxin, and FBG domain proteins) were not detected in the venom (Table 2.3Table 2.4). This included the two most highly expressed forms

of two families in the venom. Interestingly, all homologues of the boundless β-hairpin, EGF-like, K_V type 4 and 5, U_1- to U_8-Std, U_{10}-Std, and U_{12}-Std toxins were identified in our proteomic analyses.

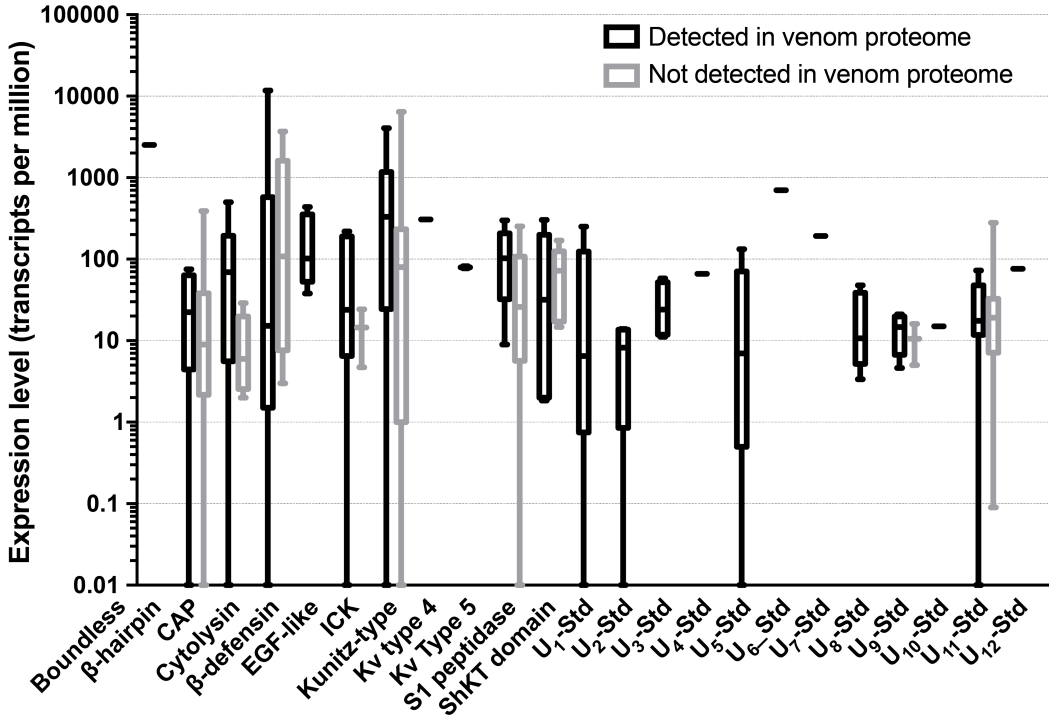


Figure 2.4 – Transcript expression levels for putative toxins detected in, or absent from, *S. haddoni* venom. Expression levels ('transcripts per million') are shown as box plots for each toxin family.

Table 2.4 – Toxin homologues not detected in the venom proteome. Ten most highly expressed putative toxins with *homology* to sequences identified in the venom of *S. haddoni*. Expression levels are shown as 'transcripts per million' (TPM) for each contig. The accession code and E-value are given for the best UniProt BLAST hit.

Transcript ID	*TPM*	*UniProt hit*	*Family*	*E-value*
TR139525_c0_g5_i1_CDS4	11699	P0DMX5	β-defensin	8E-20
TR121578_c2_g1_i1_CDS2	6421	C0HJF3	Kunitz-type	3E-16
TR90837_c2_g2_i2_CDS3	3681	P0DMX5OS	β-defensin	3E-20
TR90837_c2_g2_i1_CDS3	1788	P0DMX5	β-defensin	8E-20
TR90837_c2_g3_i1_CDS3	1075	P0DMX5	β-defensin	3E-19
TR88713_c0_g1_i2_CDS15	389	P35778	CAP	1E-13
TR145129_c1_g1_i2_CDS8	295	P0DN15	Kunitz-type	6E-32

TR105531_c0_g1_i1_CDS2	253	Q58L93	S1 peptidase	5E-41
TR145129_c1_g1_i1_CDS8	243	P0DN15	Kunitz-type	8E-35
TR124024_c0_g3_i1_CDS4	237	A6MFK8	S1 peptidase	2E-35

2.4 Discussion

This study has provided the first holistic overview of sea anemone venom. In contrast to previous work that employed an exclusively transcriptomic approach to identify toxin-like sequences (170, 177), the current study highlights the importance of combing this information with proteomic data to access the complete venom composition. Homology-based annotation of toxins from transcriptomic and genomic data is insufficient for providing a full picture of the venom arsenal (219) as it is inherently limited to finding candidates with significant sequence homology to known toxins, and it is prone to false positives because toxins evolve from housekeeping proteins and will often show significant homology to conserved ancestral sequences. Proteomics data is therefore essential to determine which proteins can be reliably identified as toxins.

The importance of incorporating proteomic data is evident from the mismatch in both the number and types of putative toxin families identified from our homology-based annotation versus the combined transcriptomic/proteomic data. The only previous example of a combined proteomic/transcriptomic approach to study sea anemone venom was reported by Rodríguez and colleagues (168). However, this study focussed on the mass fingerprint of the neurotoxic fraction and did not detail other classes of proteins present in the venom. This is the first time, to our knowledge, that deep proteomics of sea anemone venom has been used in combination with transcriptomics to identify venom constituents. Remarkably, this approach led to the discovery of 12 entirely novel protein families in the venom of a species belonging to a genus whose venom has been studied for over a century.

As mentioned above, identifying putative toxins from sequence homology is complicated by the fact that toxins typically evolve from proteins with functions unrelated to envenomation (215). This is particularly problematic for cnidarians because transcriptome data obtained from tentacles or whole body contains transcripts from tissues not specifically involved in venom production. One approach to circumvent this problem is to use gene family size to identify toxin sequences based on the rationale that large gene families result from the positive selection

by which toxins are initially thought to evolve (212, 215, 220). Transcript expression levels could also potentially be used to identify putative toxins based on the observation that toxin genes are more highly expressed in venom glands that are actively engaged in the process of venom regeneration compared to those that are replete (221).

Application of these two approaches to determine the composition of *S. haddoni* venom yielded substantially different predicted venom profiles. Based on toxin family size, the venom of *S. haddoni* is primarily composed of proteins with enzymatic activity. In contrast, based on transcript expression levels, the venom is mainly composed of peptides with neurotoxic activity. Since mRNA was extracted from tentacles engaged in venom regeneration, we would expect transcripts encoding venom constituents to be highly expressed. Indeed, this hypothesis is supported by our proteomic data, which indicate that neurotoxic peptides constitute the dominant fraction of *S. haddoni* venom. However, in contrast with previous studies of snakes and venomous arthropods (222), we did not find that expression level was a reliable indicator of which members of each toxin family are actually present in the venom. Although the detection limits of our proteomic methods could be a potential explanation, we detected neurotoxins expressed at very low levels in the milked venom (Figure 2.4). The lack of a discrete, dissectible venom-producing tissue is therefore one possible explanation for the apparent conflict between our findings and previous work (222). While our transcriptomic data was based on extracts from tentacles only, nematocytes are not restricted to sea anemone tentacles but rather are found throughout ectoderm-associated organs, some of which are known to have different toxin expression profiles

Another potential explanation for the discrepancy between detected venom components and their expression levels could be related to the process of venom regeneration in sea anemones, which remains largely uncharacterized. Surprisingly, only one RNAseq-based study of sea anemone venoms has so far specifically selected specimens actively engaged in venom regeneration (170), while none have considered the effect of regenerative states when assessing putative toxin expression levels. In fact, although the morphogenesis of nematocysts is well described (12, 223), the regeneration of sea anemone toxins is not. This study is the first to consider the regenerative state of the sea anemone venom system, and we obtained preliminary data (not shown) indicating that these cnidarians are actively engaged in toxin regeneration 72 h after venom by electrostimulation. However, further studies are required to more fully define the dynamics of toxin-gene expression after venom depletion. It is also important to point out that although our data did not include biological or technical replicates, this is also the case for

the majority of NGS-based venomics studies. Our data thus present a tale of caution, and strongly suggest that the commonly employed search-based and NGS-based gene expression level estimation approaches to identify sea anemone toxins from NGS-sequenced transcriptomic data alone may not provide an accurate reflection of the venom profile.

A large number of housekeeping proteins were identified in this study, consistent with previous work on cnidarians (178, 224, 225) This is because venom extraction via electrical stimulation is likely to induce nematocyst rupture, leading to release of cellular contents into the medium. Despite this downside of this method, it is still likely to be one of the "cleanest" methodologies for obtaining a complete venom profile. While nematocyst purification (53, 54) excludes venom components secreted from gland cells, other methodologies such as whole-tissue homogenisation or immersion in distilled water or alcohol leads to massive contamination with body proteins not associated with venom due to large-scale cell lysis (163). Although a common strategy used to eliminate structural proteins is to require the presence of a signal peptide, all structural nematocyst components, such as minicollagens and nematogalectins, carry signal peptides with similar motifs to those of toxin transcripts (12, 167, 226). Minicollagens in particular also contain cysteine-rich domains with cysteine-patterns that are reminiscent of disulfide-rich peptide toxins (12). However, minicollagens also contain characteristic Gly-X-Y repeats, which were absent in all novel putative toxins identified in this study. Moreover, nematocyst structural proteins have been characterised in detail in the starlet sea anemone *Nematostella vectensis* (53), and any that remain after our centrifugation and filtration steps are likely to be sufficiently conserved for identification by BLAST annotation.

Among the proteins we identified in the venom of *S. haddoni*, twelve families could not be assigned a putative function or protein/peptide family as they have not been previously described from any other animal venom. Even though some of the unknown putative toxin families are relatively highly expressed, they could not be detected without including proteomic evidence. Moreover, some of these families, such as U_{11}-Std, are among the main components of the venom of *S. haddoni*. This highlights the power of combining transcriptomic data with proteomic techniques not just for detecting novel bioactive proteins and peptides, but also for contributing towards an understanding of the function and evolution of venoms through a more complete description of their contents.

2.5 Conclusions

This study represents the first holistic approach to characterising the venom arsenal of a sea anemone, and it highlights how little we still know about sea anemone venoms despite decades of research. The discovery of 12 entirely new protein families in a genus of anemones that have been studied for over a century underscores the power of combining proteomic and transcriptomic data when investigating animal venoms. Although there is still much to learn about the composition of sea anemone venoms and the role of individual venom components in prey capture, defence and intraspecific competition, this work provides a solid foundation for future research into the ecology and evolution of these venoms (219). However, much work will be required to understand the function of individual venom proteins and peptides, and how they contribute to the overall envenomation process.

CHAPTER 3
Evolution of the sea anemone venom arsenal

Abstract

Venoms are among nature's most complex cocktails that are characterised by a diversity of molecules, such as large proteins, small peptides, polyamines, and salts, which disrupt the physiology of prey animals upon injection. Several gene families encoding venom components have been subjected to extensive duplication and have evolved under the influence of positive (diversifying) selection. The phylum Cnidaria (sea anemones, jellyfish, corals, sea pens and hydroids) is the oldest known venomous animal lineage (~750 million years old). More than a century of research on sea anemone venoms as a potential source of novel bioactive therapeutics has shown that they contain a diversity of biologically active proteins and peptides. However, recent omics studies have revealed that much of the venom proteome remains unexplored and that the diversification and molecular evolutionary regimes of toxins encoded by these fascinating and ancient animals remain poorly understood. Here, we present the most comprehensive insight into the composition and evolution of sea anemone venoms to date. We used combined transcriptomic and proteomic data of milked venom from five genera of sea anemones to further interrogate previously published transcriptomes from an additional eight taxonomically diverse species of sea anemone. Our results reveal more than 1500 unique toxin-like sequences which were grouped into 38 families. 20 of these 38 families had no blast homology to any known protein and peptide family. The presence of a wide range of novel proteins and peptides families discovered here highlights these animals as a rich source of novel bioactive molecules. Moreover, phylogenetic reconstruction shows fascinating insights into the evolutionary origin and diversification of sea anemone toxin.

3.1 Introduction

Venoms are among nature's most complex cocktails and are characterized by a diversity of molecules, including large proteins, small peptides, polyamines, and salts, which disrupt the physiology of prey animals upon injection (215) Because of their key roles in many inter- and intra-specific interactions between animals, venom toxins have independently evolved several times across diverse animal lineages, resulting in toxins targeting a variety of functionally important protein complexes and macromolecules involved in cellular homeostasis (215).

Cnidaria is believed to be one of the most ancient venomous eumetazoan lineages, having evolved since Neoproterozoic times, ~650 million years ago, long before the Cambrian radiation (227). There has been resurgence in interest surrounding the nature and evolutionary origins of cnidarian venom toxins because the impact on human health (228) and biotechnological potencial as source of new pharmacological tools and/or therapeutical leads (229). Because certain cnidarians serve pivotal ecological roles and are important model organisms in the field of evolutionary developmental biology, genomic and transcriptomic data for several cnidarian species have been rapidly accumulating in recent times (85, 230, 231). However, proteomic data from cnidarian venoms remains scarce. This is particularly problematic for cnidarians because transcriptome data obtained from tentacles or whole body contains transcripts from tissues not specifically involved in venom production. As shown in Chapter 2, homology-based annotation of toxins from transcriptomic and genomic data is insufficient for providing a complete picture of the venom arsenal, as it is inherently limited to finding candidates with significant sequence homology to known toxins. In addition, purely homology-based methods are prone to false positives due to the fact that toxins typically evolve from proteins with functions unrelated to envenomation.

Sea anemones (Anthozoa) are benthic, sessile cnidarians that use venom for a wide variety of functions, such as prey capture, defence, digestion, and inter- and intraspecific competition. As a result of hundreds of millions of years of evolution (232-234), and probably also the diversity of roles that venom plays, sea anemones have evolved a rich variety of biologically active compounds (148, 235). They also appear to be unique among cnidarians in having a venom that is rich in peptide toxins (54). This has made sea anemones particularly attractive to drug discovery efforts, as evident from the decades of research into the structure and function of sea anemone toxins (236). However, despite the long history of biomedical research into sea anemone venoms, there is a striking lack of non-activity guided studies on their composition.

Reflecting this, sea anemone venoms are in fact rich in peptides and proteins that bear no resemblance to any known protein and peptide family (Chapter 2). This wealth of unknown structural folds and pharmacological properties, combined with an overall poorly sampled taxonomic lineage, limits our understanding of the evolution of sea anemone toxins.

In this study, we provide the first comprehensive and detailed insight into the evolution of sea anemone venom. To access the composition of milked venom from five sea anemone species, we used a combination of transcriptomic and proteomic techniques. This approach created an accurate and comprehensive list of venom components, and we used this data to identify toxin orthologues in an additional eight sea anemone transcriptomes selected from a broad taxon sample in order to more accurately identify toxin recruitment and evolutionary events. Our results confirm previous findings of dynamic evolution in cnidarian venoms and reveal a complex picture of both ancient and lineage-specific recruitments and functional radiations among sea anemone toxins.

3.2 Results and Discussion

To identify the proteins and peptides present in milked venom of *A. pulchella*, *A. tenbrosa*, *H. malu*, *M. doorensis* and *S. haddoni*, we used a combination of transcriptomics and 2DE and shotgun analyses of crude and RP-HPLC fractionated venom (Figure 3.1, methods section for transcriptome assembly details). We used next-generation sequencing of poly-A enriched RNA extracted from tentacles that were actively regenerating their venom to investigate the diversity of toxin-like sequences in sea anemone venom. De novo assembly with Trinity yielded 157,354 contigs in *A. pulchella*, 87,485 in *A. tenebrosa*, 304,320 in *H. malu*, and 303,941 in *M. doorensis*. The contigs were translated to all possible coding sequences, resulting in 1,108,098 amino acid sequences in *A. pulchella*, 457,470 in *A. tenebrosa*, 2,141,382 in *H. malu*, and 2,296,168 in *M. doorensis*. *S. haddoni* results presented here were obtained previously on chapter 2 but included in this chapter as comparison. Our RP-HPLC fractionation and 2DE gel revealed a complex set of both low and high molecular weight proteins, and the toxin types present in the milked venom of each species are summarized in Table 3.1. The final set of potential toxins identified in the milked venoms included 18 known toxin families and 20 protein families that could not be assigned a putative function or known family, 12 of which were identified previously. These were label as U, referring to unknown, followed by a number. Moreover, the numbering order was organised according to the number of cysteines in the

mature peptide, where U1 has only 2 cysteine residues and U20 has 20. All the potential toxins were combined with all sea anemone toxins in UniProt (96) to generate a comprehensive database of confidently annotated sea anemone venom components.

Even though the present chapter is the first comprehensive insight into toxin recruitment events across a wide range of sea anemone taxa, most of the species used for proteomics analyses belong to the superfamily Actinioidea, with only one species from Metridioidea included in the study. Actinioidea is a larger group and easier to collect, which had an impact in the choice of the species. Thus, some ascertainment bias might exist in the present dataset. Another fact to be consider is the venom profile across the life cycle. The results presented in this chapter were obtained from adult animals. Columbus-Shenkar and colleagues (237) showed that different toxins can be presented in different stages of life. Unfortunately, we could not cover this topic in this study, and the venom proteome might more diverse than here reported. Although outside the scope of this thesis, this would be an interesting topic for a future study.

To investigate the full diversity of each toxin family identified in our proteomic data and also to have information of venom composition in different species, we conducted a BLAST search with the custom database against the transcriptomes of an additional eight species across the Actiniaria phylogenetic tree (see materials and methods section), as well as those used for proteomic characterisation of milked venom. Sea anemone toxins can be classified into three major functional groups, which are neurotoxins, enzymes and pore-forming toxins Figure 3.2. However, the 20 families of previously undescribed putative protein and peptide toxin families may or may not belong to one of these categories require further characterization to confirm if sea anemone venom components can still be categorised into only these three broad functional groups.

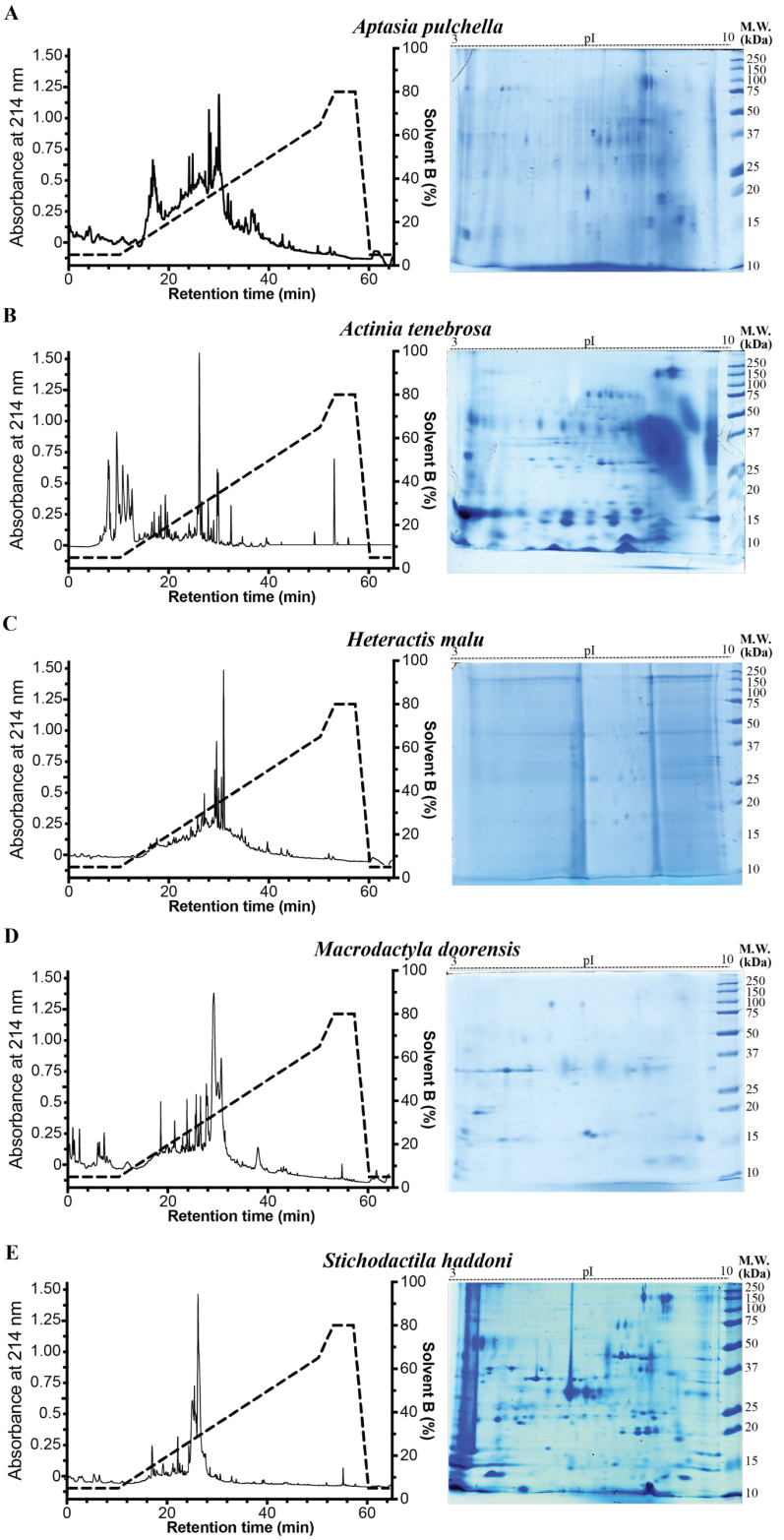


Figure 3.1 — Proteomic analyses of five sea anemone venoms. Composition of venom obtained by electrostimulation is show as rpHPLC chromatrogams (left) and 2D-PAGE gels (right) for A) *Aiptasia pulchella*, B) *Actinia tenebrosa*, C) *Heteractis malu*, and D) *Macrodactyla doorensis*, as well as E) *Stichodactila haddoni* for comparison (see Chapter 2). For 2DE gels the Isoelectric point (pI) is indicated above and molecular weight (kDa) on the right of each gel.

Table 3.1 — Summary of protein families in proteomics of sea anemone venoms. Table shows in which species in this study was identified the toxin families listed.

Family	*A. pulchella*	*A. tenebrosa*	*H. malu*	*M. doorensis*	*S. haddoni*
Acrorhagin		x			
Actinoporin		x			
BBH-like					x
Defensin-like	x	x	x		x
EGF-like		x			x
ICK-like		x		x	x
Kazal-like domain	x				
Kunitz-type		x		x	x
Kv type5		x		x	x
PLA_2		x			
SCRiP				x	
Sea Anemone 8			x		
ShK-like		x	x	x	
Peptidase S1					x
Peptidase M12					
FactorV-like	x	x			
U1		x	x		
U2		x	x		
U3 (U_1Shd)					x
U4	x				
U5 (U_1Shd)					x
U6	x				
U7 (U_4Shd)	x				x
U8 (U_5Shd)					x
U9 (U_6Shd)	x				x
U10			x		
U11 ($U_{7\&8}$Shd)					x
U12				x	
U13				x	
U14		x		x	
U15 (U_9Shd)					x
U16		x	x	x	
U17				x	
U18 (U_{11}Shd)					x
U19	x				
U20 (U_{12}Shd)					x

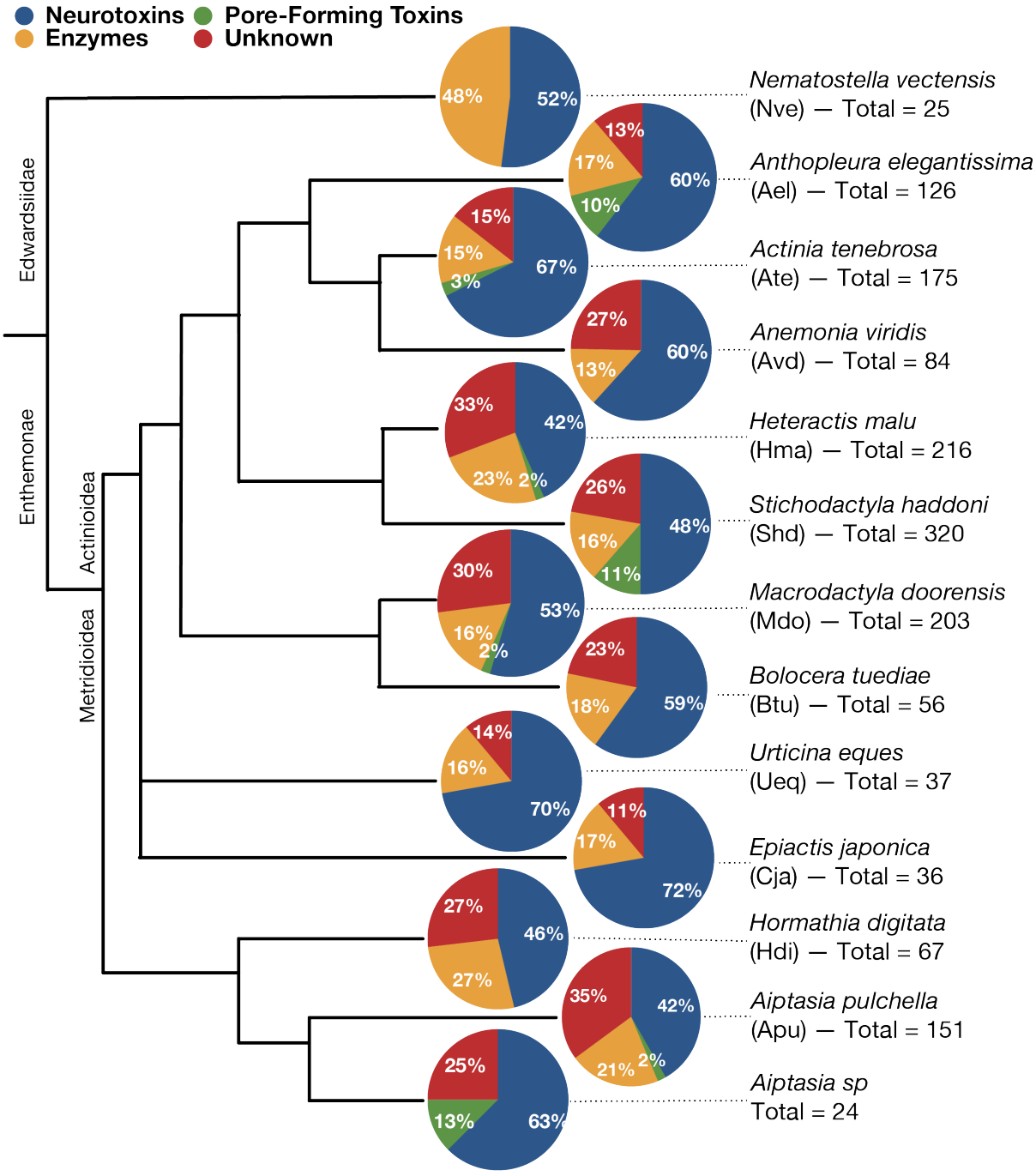


Figure 3.2 – Putative broad functional venom profiles of sea anemones. Putative toxins identified in transcriptomes of sea anemones based on our proteomics database. Putative neurotoxins are cysteine-rich venom peptides related to known neurotoxic sea anemone. Total number of homologues identified is shown.

Neurotoxins are the major component in sea anemone venoms (Figure 3.2). This is perhaps not surprising given the predominantly sessile natural history of sea anemones, most of them spend the adult life attached to a surface, such as rocks and/or sediment. Because of that, they rely on

the neurotoxic effect of their venom to paralyse and kill prey and also for defence from predators (237). Interestingly, the venom of *Nematostella vectensis*, sister to other sea anemone groups, appears to be just composed by enzymes and neurotoxins, and entirely lacking in cytolysins (Figure 3.2). Furthermore, this is one of the species with the least number of toxins and putative toxin orthologues identified, and we did not find any representatives of any of the new unknown putative toxins. In contrast, a substantial fraction of the venom components (13-35%) in the other species are completely novel (Figure 3.2).

The sea anemones with greatest number of toxins identified were *S. haddoni* (320 unique sequences), *H. malu* (216 unique sequences) and *M. doorensis* (203 unique sequences) (Figure 3.2). Curiously, these sea anemones have similar ecology and live in association with anemonefishes. For establishment and maintenance of this symbiotic relationship, sea anemones have to offer potential fitness benefits to the anemonefish and also the other way around (238). Anemonefish defends the anemone from its predators and parasites (239). Also, sea anemones can benefit of nutrients from the anemonefish's excrement (240). In the other hand, sea anemones offer protection to anemonefishes by living within the stinging tentacles (241). If toxicity of its venom is too low, anemonefish will not be able to obtain the benefits from the association. Moreover, to maximize fitness, anemonefish should choose anemone hosts that provide them with the highest quality refuge at the lowest cost to themselves with respect to physiological expenditure. These factors could bring extra complexity to the venom arsenal of these sea anemones.

Another factor for the venom of *S. haddoni*, *H. malu* and *M. doorensis* be more diverse than the other species studied in this chapter could be the lack of behavioural defence. High predation pressure in the marine environment brings strong selective pressure on sessile or slow-moving organisms to evolve defence mechanisms. Chemical defence, behavioural defence, and a symbiotic partner that provides protection are the main defence strategies which sea anemones possess. These strategies can or may not be mutually exclusive. Usually, there is an inverse correlation between chemical and behavioural defence. Sea anemones with the ability to defend themselves by completely burrow into the sediment do not rely as much on the chemical defence as the ones that do not present this behaviour. In this case, the higher number of toxins identified in *S. haddoni*, *H. malu* and *M. doorensis* may be sufficient as a defence strategy against predators so that withdrawal under the sediment is not required.

3.2.1 Neurotoxins

As mentioned above, the putative toxins with the greatest number of homologues identified by BLAST searches were neurotoxins. The final set of potential neurotoxins identified across all species studied here were Acrorhagin (28 homologues), ATXIII (7 homologues), BBH-like (53 homologues), β-defensin-like (151 homologues), EGF-like (16 homologues), ICK-like (194 homologues), Kazal-like (17 homologues), Kunitz-like (177 homologues), K_V type 5 sea anemone toxins (12 homologues), SCRiP (30 homologues), Sea anemone 8 (81 homologues) and Shk-like (56 homologues) (Figure 3.3). In total, more than eight hundred unique CDSs of neurotoxin-like peptides were identified.

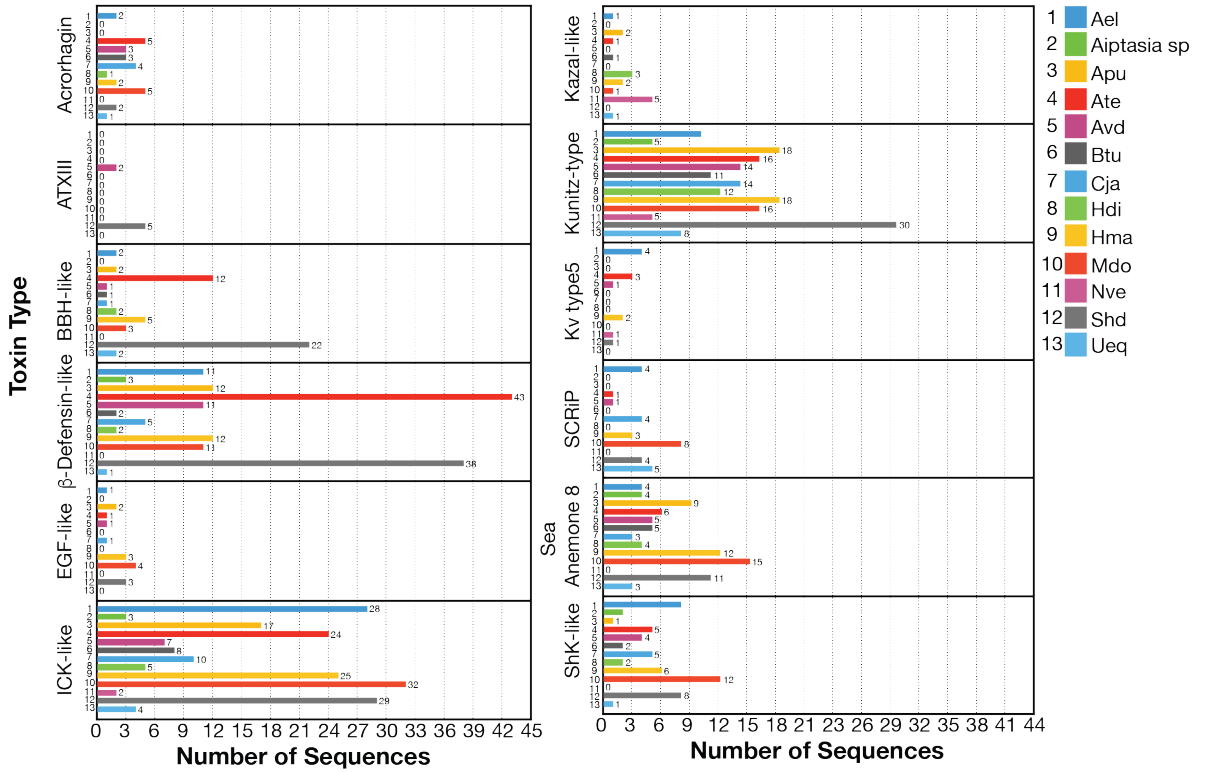


Figure 3.3 — Venom neurotoxin scaffold (family) diversity across sea anemones. Function annotation of the unique CDSs identified by BLAST search using our proteomics database. Identical sequences were removed using CD-HIT. Species were abbreviated according to Oliveira and colleagues (33). Each species is indicated by a number and colour as shown in the key at right side.

To have insights about toxin recruitment, we summarise the phylogenetic distribution of the neurotoxins in Figure 3.4. ICK-like, Kunitz-like and β-Defensin-like were recruited early in evolution and all species studied here have these scaffolds in their venom. Five neurotoxin types were identified in the venom of the suborder Edwardsiidae (*N. vectensis*): ICK-like, Kunitz-like, β-Defensin-like, Kazal-like, and K_V type 5 sea anemone toxins. Sea anemone

family 8 and ShK-like were recruited early in the evolution of Enthemonae suborder. Acrorhagin and BBH-like also could have appeared early in the evolution of Enthemonae suborder but Acrorhagin was lost in the Aiptasiidae family and BBH-like in some of its species. So far only four Acrorhagin-like toxins were reported to be presented in sea anemone venom (169). They were discovered in the acrorhagi of the same species. However, here we found Acrorhagin-like toxins in species which do not have acrorhagi, suggesting this family is not exclusive in Acrorhagi nematocytes. SCRiP was just recruited by a branch of the family Actiniidae. A secondary loss of Kv type 5 and BBH-like is likely to happen in the general *Bolocera*, *Urticina*, and *Epiactis*. Similar event apparently happened for EGF-like for some species in Enthemonae suborder. ATXIII homologues were just found in two species, both from the superfamily Actinioidea (*A. viridis* and *S. haddoni*), suggesting this is a recent recruitment. Below the three most abundant scaffolds will be described in more details.

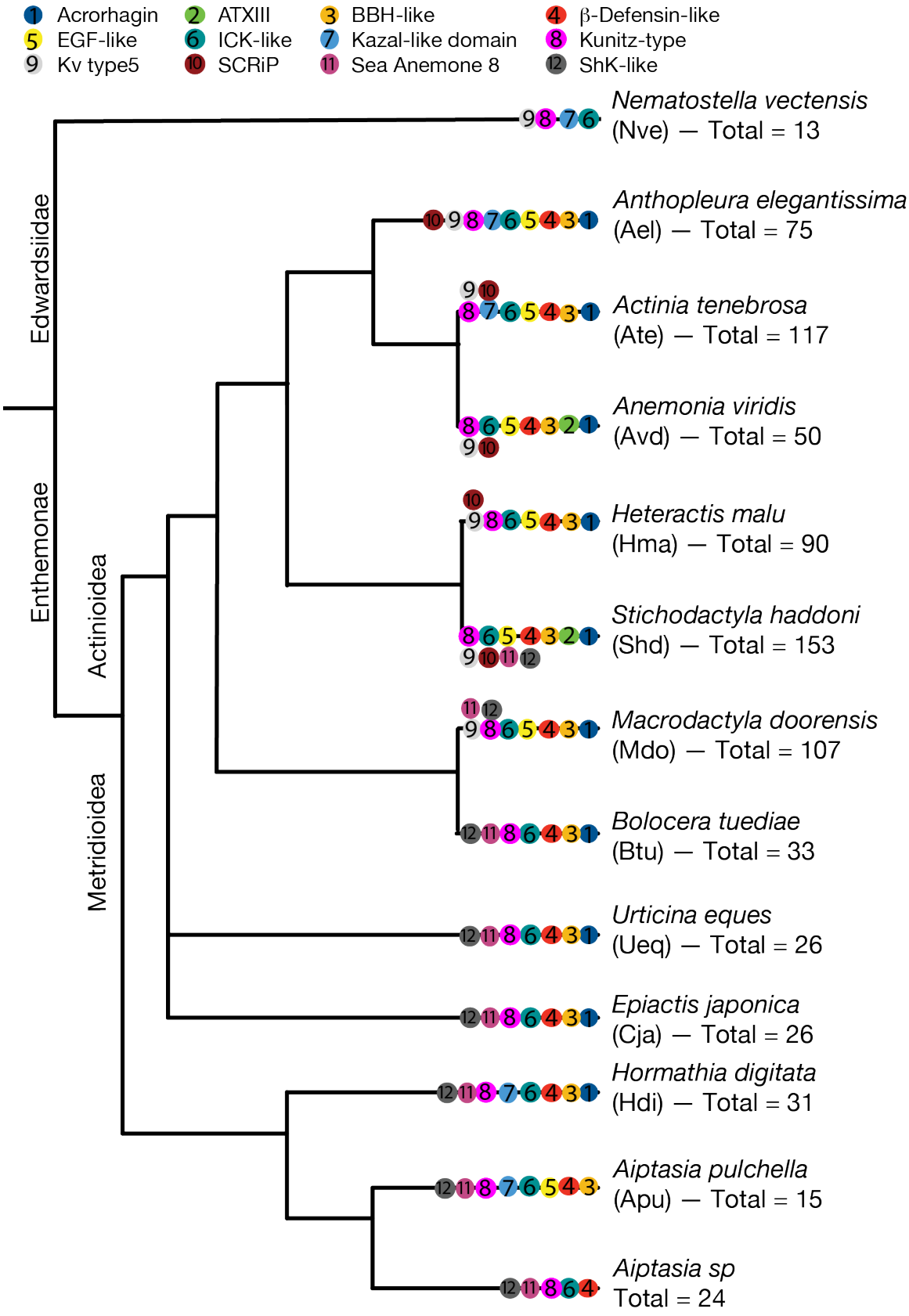


Figure 3.4 — Phylogenetic distribution of neurotoxins families in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of neurotoxin families as their earliest respective recruitments.

3.2.1.1 ICK-like

The ICK fold is probably the most widely recruited peptide-fold in animal venoms. This frequent "weaponization" is probably attributable to both the stability and evolutionary plasticity of the ICK fold (242). ICK toxins are particularly diverse in spider venoms but this scaffold is also present in the venoms of scorpions (243), assassin bugs (244), cone snails (245), tick saliva (246). Moreover, putative ICK toxins have been identified in the venoms of ants (247), remipedes (248) and centipedes (249). Although the 3D structure of sea anemone ICK peptides remain to be confirmed, a toxin displaying cysteine pattern characteristic of the ICK structural scaffold, namely PhcrTx1 (π-phymatoxin-Pcf1a) were reported (88). This toxin reversibly inhibits acid-sensing ion channels (ASIC) in rat dorsal root ganglia neurons (IC_{50}=100 nM).

In the present work, we found 194 unique sequences with classic cysteine pattern characteristic of the ICK signature. Surprisingly, ICK motive is present in the venom of all species analysed and it is the scaffold with greatest number of homologues identified. Unfortunately, not enough data is available to gain insights of the functional evolution of this structural motif in sea anemone venoms. However, the diversity of sea anemone ICKs suggests they interact with a range of molecular targets (Figure 3.5). It will be interesting to search for ICK motif outside the order Actiniaria to examine its prevalence and evolution in Cnidarians.

In view of the fact that K_V type 5 sea anemone toxins has similar cysteine framework as ICK with an extra disulfide bond (242) (C-C-CC-C-C-C-C), we included these in our phylogenetic analyses (Figure 3.5). K_V type 5 form a distinct clade and is likely evolved from a classic ICK signature.

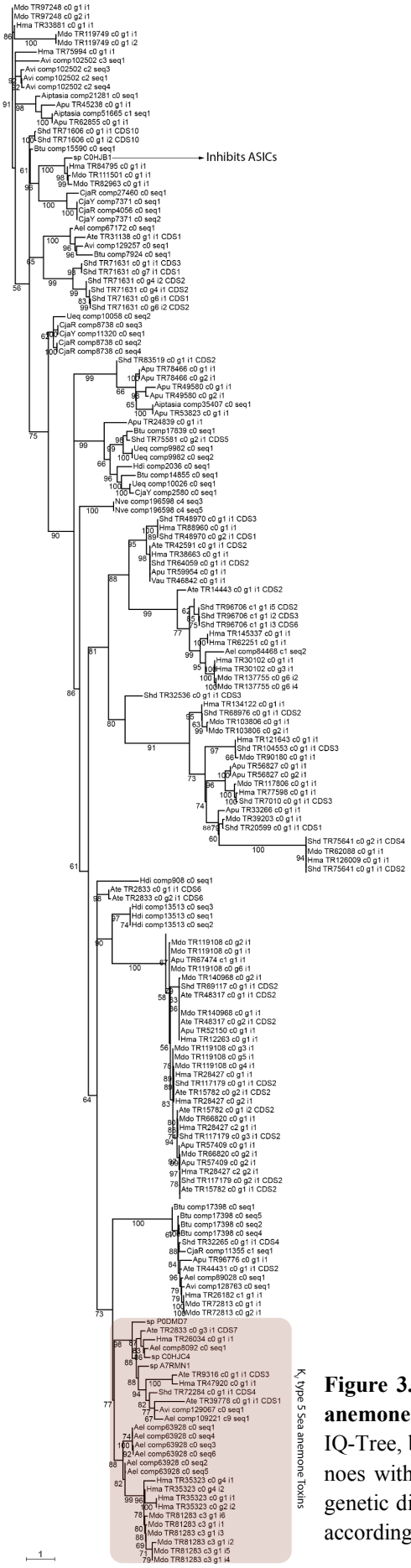


Figure 3.5 — Phylogenetic reconstruction of ICK homologues in sea anemone venoms. Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+I+G4 model, which was the best fit model according to Bayesian information criterion (BIC).

3.2.1.2 Kunitz-like

Sea anemone toxins with Kunitz-type motif consists of peptides of ~60 amino acid residues stabilized by three disulfide bridges (C_1–C_6, C_2–C_4, C_3–C_5). Its 3D structure is characterized by an α/β/α motif with a hydrophobic core (144). The peptides of this family have an ancient Kunitz fold and some of them are characterized by a unique and intriguing feature of dual functionality since they inhibit both proteases and ion channels (27, 250). It has been hypothesised that Kunitz peptides in sea anemone venom protect toxins from protease degradation during storage but are also paralyze prey through inhibition of ion channels (214). Supporting the notion that Kunitz-like toxins are important constituents of sea anemone venoms, this is the second most abundant structural motif recruited among the sea anemone neurotoxins, with 177 unique sequences identified.

Phylogenetic analysis of the Kunitz-like toxins found here in addition to the ones already submitted at Uniprot database, suggests this family was recruited prior to the split of Enthemonae and Edwardsiidae (Figure 3.4). In fact, a recent study discovered kunitz-like peptides in the zoanthid *Palythoa caribaeorum* through transcriptome sequencing (251). This suggests an early recruitment in the subclass Hexacorallia, however the lack of research on octocorallians venoms prevents us from determining if Kunitz-like toxins were recruited before Anthozoans split in these subclasses.

Since the Kunitz-like toxins appeared in sea anemones, substantial diversification happened. Phylogenetic analysis (Figure 3.6) suggests this family original function was protease inhibitor of trypsin. This is also a known activity for non-toxin Kunitz-peptides (214). Later on, it diversified its function to inhibit other proteases (e.g. serine, cysteine, and aspartic proteinases) and the function to inhibit ion channels was acquired. After the ability to block K_V1 was recruited, it was followed by explosive diversification. This confirms that blocking K_V channels is a successful strategy for prey capture and defence in sea anemone venoms. Even though Kunitz-like peptides are a well-studied sea anemone toxin family, our analysis shows there is a large clade of functionally uncharacterised peptides in this family, demonstrating the discovery potential even among peptides in relatively well characterised toxin families.

Figure 3.6 — Phylogenetic reconstruction of Kunitz-like peptides in sea anemone venoms. Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+R5 model, which was the best fit model according to BIC.

3.2.1.3 β-Defensin-like

β-Defensins are antimicrobial peptides that are secreted as part of the innate immune response in a wide range of taxa (101, 102). However, in sea anemone venoms, β-defensin-like peptides have become weaponized to serve as neurotoxins that modify the activity of voltage- and ligand-gated ion channels; this family of peptides includes K_V type 3, Na_V type 1, 2 and 4 and ASIC toxins (103-106). The pharmacological diversity in this neurotoxic lineage is probably reflective of an adaptive radiation a and also due the early recruitment. This scaffold is present in all sea anemones venoms studied here and it is the third most abundant peptide motif, with 151 unique sequences identified. It is important to also note that although there are more than a hundred sequences deposit at Uniprot database, the target of the vast majority (56 out of 102 toxins) was assumed by similarity with other toxins. This is a distinction that is critical to be able to accurately infer the evolutionary history of this family. Nevertheless, phylogenetic analysis of this family suggests K_V blockers and Na_V modulators are predominantly found in different clades (Figure 3.7). These findings differ from those of Jouiaei and colleagues (162), who proposed that a subset of sodium channel targeting toxins experienced episodic bursts of adaptive selection in Actinioidea and accumulated mutations at an elevated rate, resulting in the origination of a novel toxin type that can target potassium ion channels. A critical factor was they failed to find homologues of β-defensins which blocks K_V channels in sea anemone transcriptomes outside Actinioidea (*Nematostella vectensis* (252), *Edwardsiella lineate* (253), *Metridium senile* and *A. pallida* (254)). In contrast, we found in three non-actinoidean species.

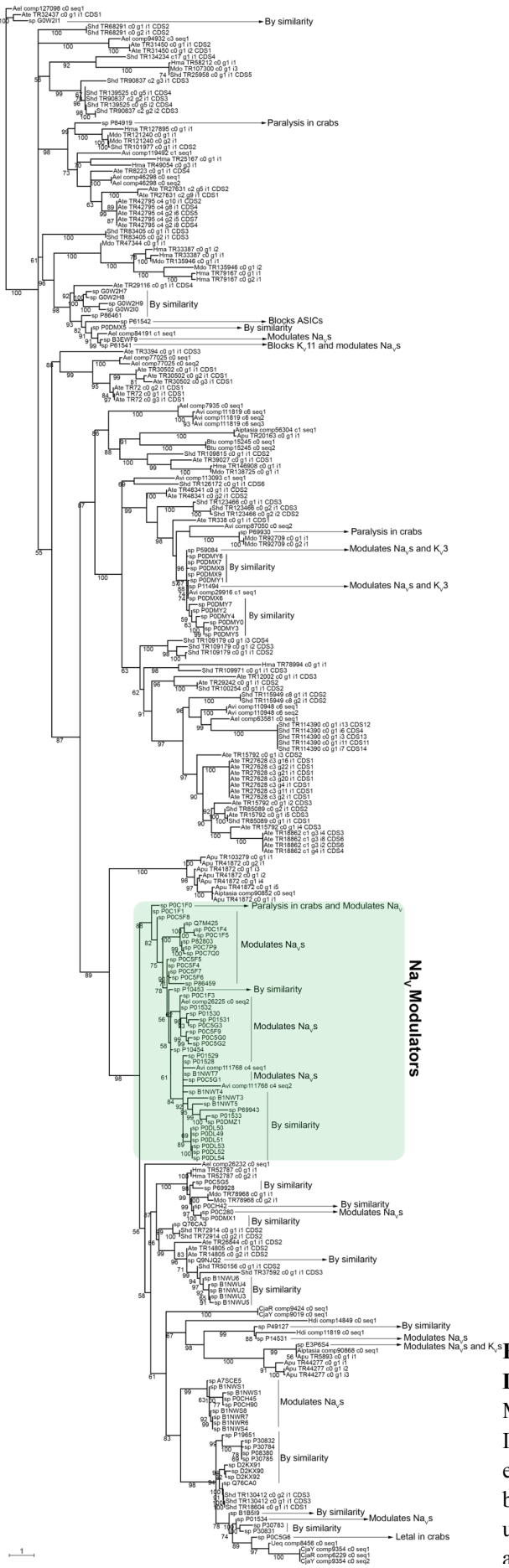


Figure 3.7 — Phylogenetic reconstruction of β-Defensin-like peptides in sea anemone venoms. Maximum-likelihood unrooted tree was calculated using IQ-Tree, bootstrap support values above 50 are shown at each node, while noes with support of 50 or less have been collapsed. Scale bar indicates genetic distance under the VT+R5 model, which was the best fit model according to BIC

3.2.2 Pore-forming toxins

Cytolysins are a group of sea anemone toxins that form pores in cell membranes, and therefore belong to a larger group of 'pore forming toxins' (PFTs) (44). So far, sea anemone PFTs can be classified in five types. We only found actinoporins in our proteomics data (type II). Actinoporins are one of the better characterized PFTs from sea anemone venoms. These proteins form a pore in cellular membranes containing sphingomyelin. Actinoporins are comprised of a single domain (~20 kDa), lack cysteine residues, and are equipped with functionally important regions conserved throughout the toxin gene family (50, 59). Our finds suggest actinoporin homologues are only present in the venom of Actinioidea (Figure 3.8 and Figure 3.9). Recently, the evolution and identification of possible functionally important residues of actinoporins were extensively studied by Macrander & Daly (255). Our results confirm their finding that no candidate actinoporins were present in either of the two examined edwardsioideans (*Nematostella vectensis* and *Edwardsia elegans*) (255). On the other hand, actinoporin-like toxins have been described in only one non-actiniarian species, *Hydra magnipapillata* (256), making the evolutionary processes of actinoporin further complicated.

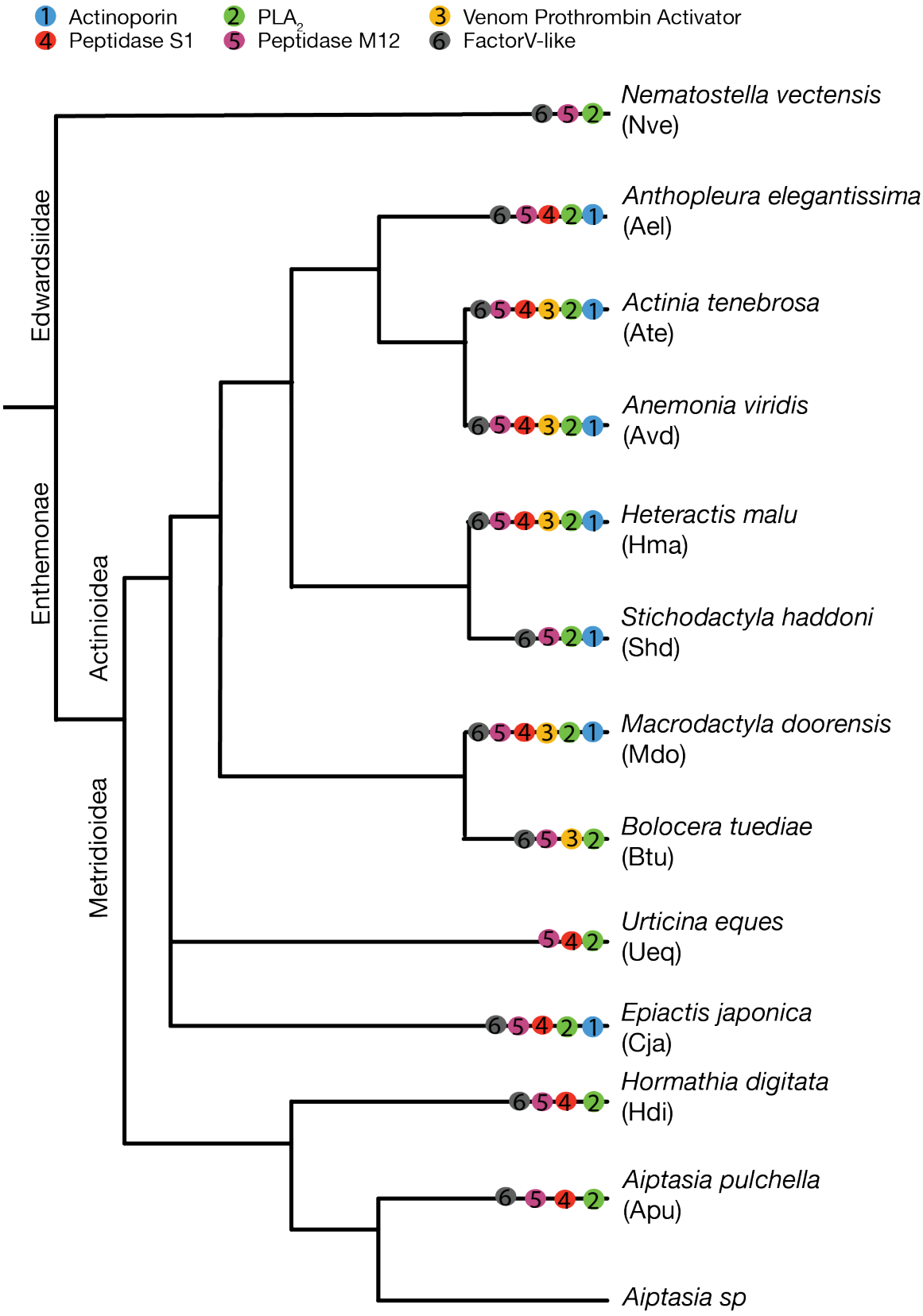


Figure 3.8 — Phylogenetic distribution of actinoporin and enzymes families in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of actinoporin and enzyme families as their earliest respective recruitments.

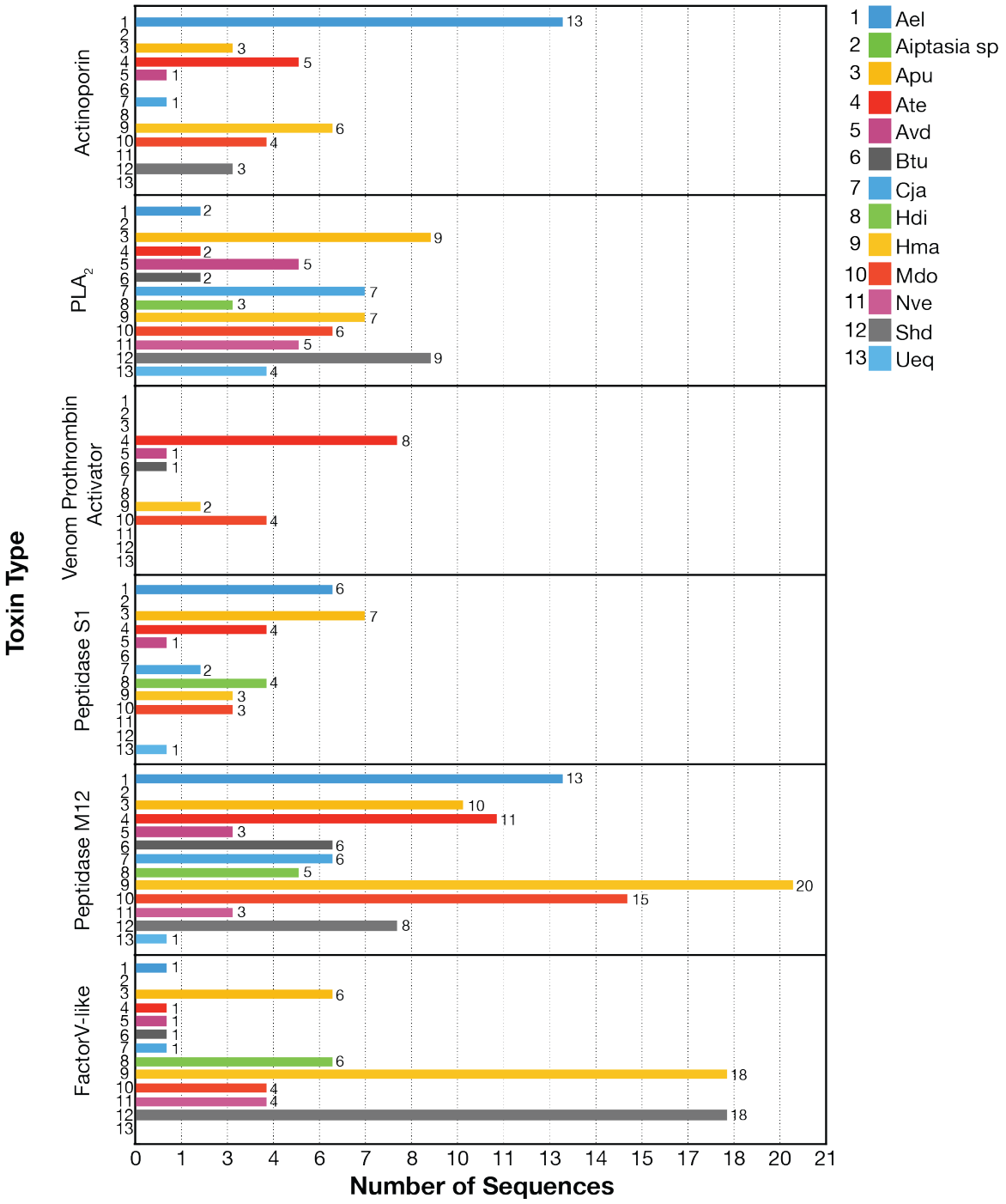


Figure 3.9 — Venom pore-forming toxin and enzymes (family) diversity across sea anemones.

3.2.3 Enzymes

It is believed that enzymes play a major hole in sea anemone venoms to help in defence, prey capture and digestion (148). However, there is a large discrepancy between the types of

enzymes reported from transcriptomic studies of sea anemones with those reported from studies on milked venom. As previously mentioned, sea anemones do not have a centralised venom gland, and it is therefore difficult to use transcriptomic techniques to distinguish between enzymes that have housekeeping roles and those that play a role in envenomation. This makes the present work important to identify which families of enzymes are actually venom constituent. Five families of enzymes were found in our proteomics data and also across the species in analyses (Figure 3.9). They are PLA_2, venom prothrombin activator, peptidase S1, peptidase M12 and factor V-like. PLA_2, factor V-like, peptidase S1 and M12 are widely distributed, being present in both Enthemonae and Edwardsiidae suborder (Figure 3.8). This suggests an early recruitment before these suborders split. Venom prothrombin activator was found only in Actinoidea. Unfortunately, activity data is really scarce to give support for further phylogenetic analyses.

3.2.4 Venom components without BLAST homology

In total, 399 unique sequences with no blast homology were found (Figure 3.10). These were organised in 20 families, many of which were both abundant and diverse. All of these novel peptide and protein families are cysteine-rich (Figure 3.12Figure 3.13). This highlights the power of combining transcriptomic data with proteomic techniques not just for detecting novel bioactive proteins and peptides, but also for contributing towards an understanding of the function and evolution of venoms through a more complete description of their contents. The 3 most abundant families were U2, U11 and U15, with 41, 45 and 48 unique sequences identified across all species, respectively (Figure 3.10). These families are present in almost all species of Enthemonae suborder, suggesting they share an ancestor. Characterization of these new families is essential for better classification and further phylogenetic analysis.

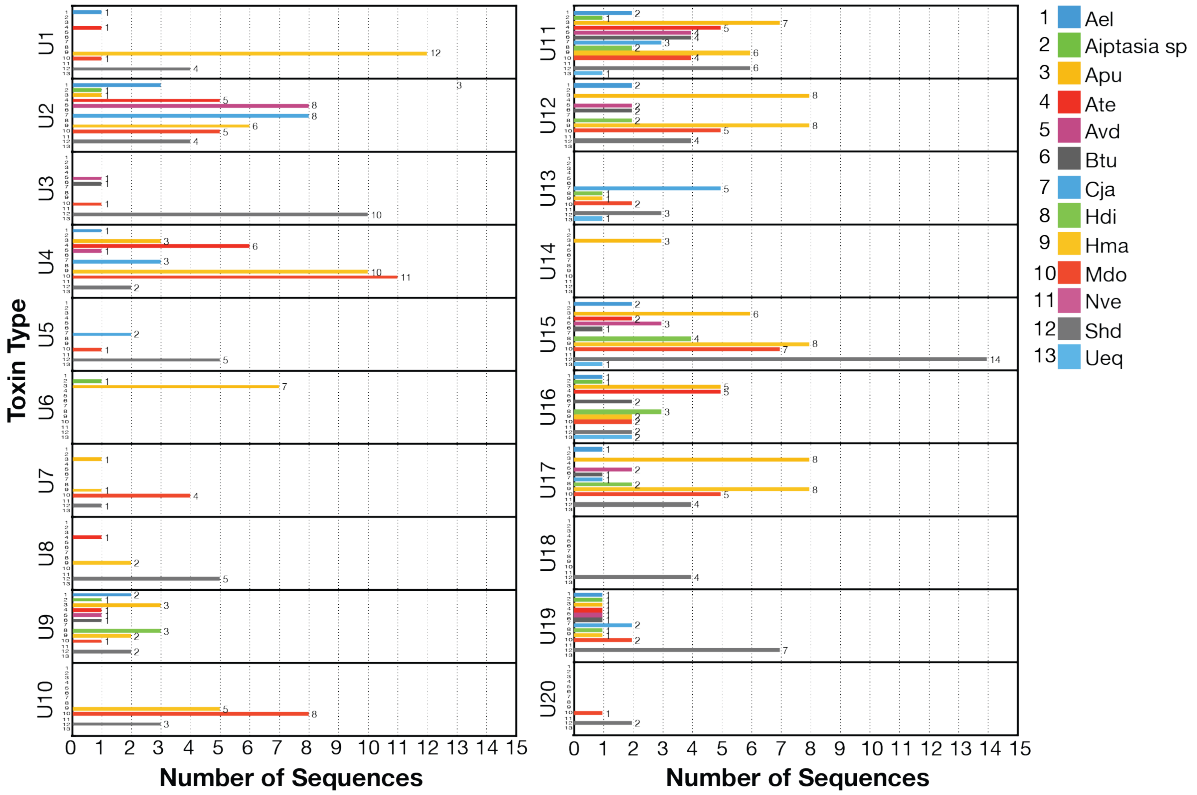


Figure 3.10 — Diversity of proteins with no blast homology across sea anemones.

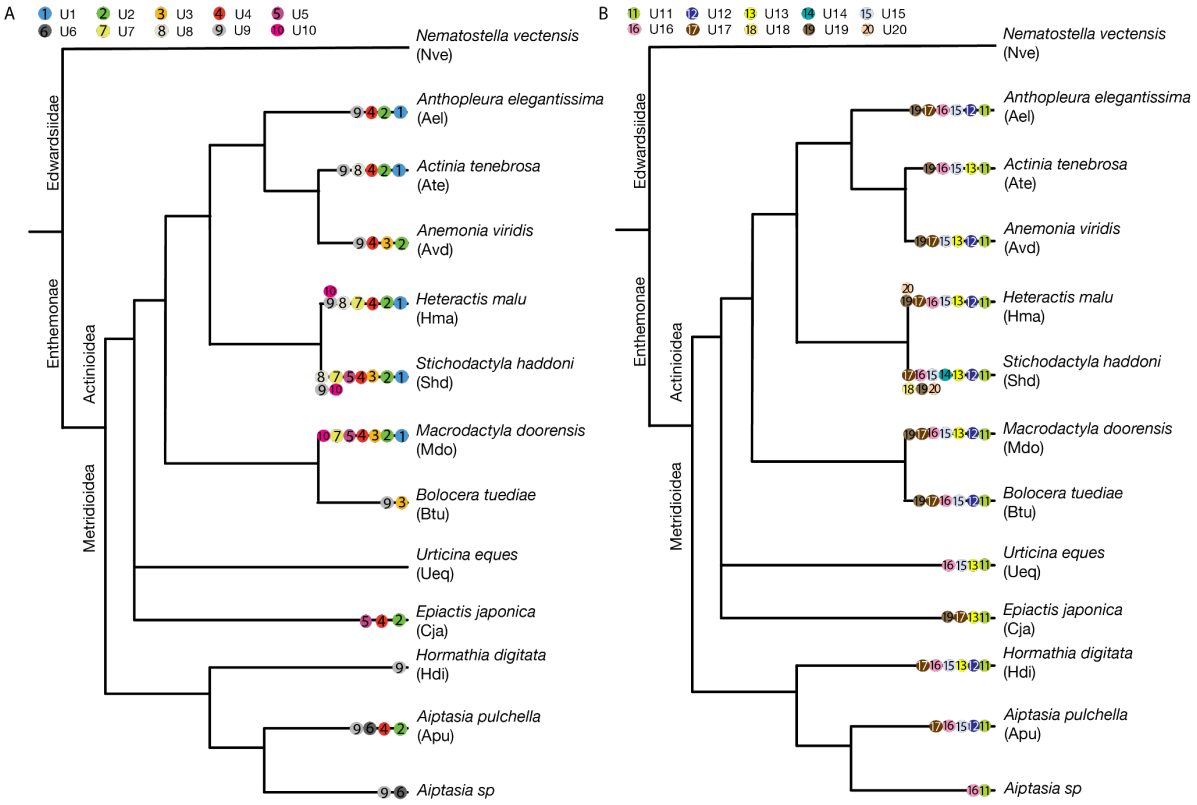


Figure 3.11 — Phylogenetic distribution of proteins with no blast homologies in sea anemones. A representative phylogenetic tree of the species included in this study, showing the phylogenetic distribution of these novel families. **A** Families U1 to U10, and in **B** Families from U11 to U20.

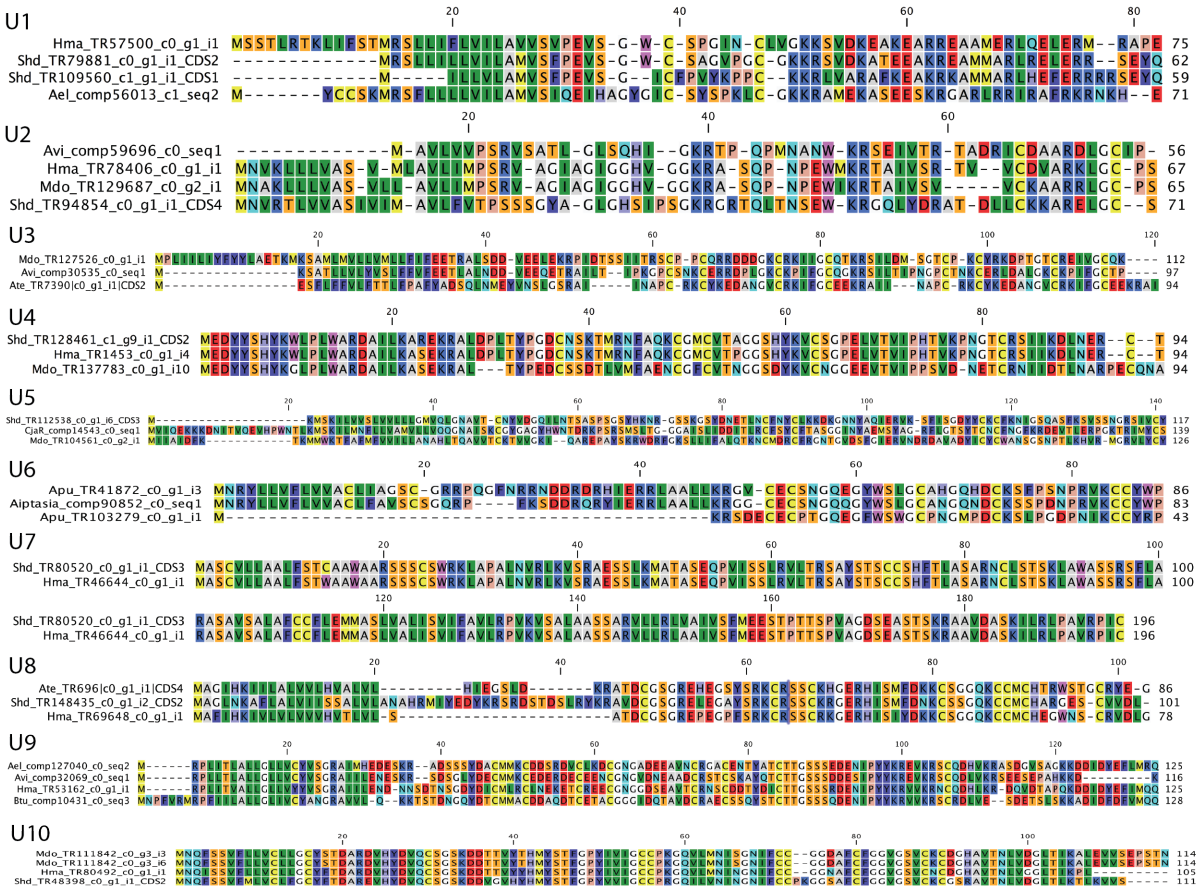


Figure 3.12 — Representative sequence of proteins with no blast homology. Alignment of representative sequences of U1 to U10.

Figure 3.13 — Representative sequence of proteins with no blast homology. Alignment of representative sequences of U11 to U20.

3.3 Conclusion

This study has provided the first holistic overview of sea anemone venom composition and evolution. In contrast to previous work that employed an exclusively transcriptomic approach to identify toxin-like sequences, the current study highlights the importance of combing this information with proteomic data to access the complete venom composition. The importance of incorporating proteomic data is evident from the number of families with discovered. This approach led to the discovery of 20 entirely novel protein families and homologues in different species.

Among the proteins we identified in the venom proteomics, twenty families could not be assigned a putative function or protein/peptide family as they have not been previously described from any other animal venom. Even though some of the unknown putative toxin families are relatively highly abundant and diverse, they could not be detected without including proteomic evidence. Moreover, some of these families, such as U11, are present in most of the venoms studied here. This highlights the power of combining transcriptomic data with proteomic techniques not just for detecting novel bioactive proteins and peptides, but also for contributing towards an understanding of the function and evolution of venoms through a more complete description of their contents.

This study represents the first holistic approach to characterising the venom arsenal of a sea anemone, and it highlights how little we still know about sea anemone venoms despite decades of research. The discovery of 20 entirely new protein families underscores the power of combining proteomic and transcriptomic data when investigating animal venoms. Although there is still much to learn about the composition of sea anemone venoms and the role of individual venom components in prey capture, defence and intraspecific competition, this work provides a solid foundation for future research into the ecology and evolution of these venoms (219). However, much work will be required to understand the function of individual venom proteins and peptides, and how they contribute to the overall envenomation processes. Hopefully, our contribution will help spark a greater interest in ecological and evolutionary aspect of these venoms.

3.4 Materials and methods

3.4.1 Specimen and Venom Collection

Sea anemones were collected at North Stradbroke Island, Queensland, Australia (27°15 S, 153°15 E) then housed in aquaria at the University of Queensland. Venom was obtained after a starvation period of at least 48 h. Briefly, the sea anemone was rinsed, placed in a minimal volume of artificial seawater, and the nematocysts induced to discharge by electrical stimulation (179). The water, which contains the venom, was lyophilized and then the venom was desalted by dialysis at 4°C (Biotech Cellulose Ester membrane, 0.1–0.5 kDa cut-off; Spectrum Labs, USA).

3.4.2 2DE Analysis

Desalted venom (0.4 mg) was solubilized in 125 µL of DeStreak Rehydration Solution (GE Life Sciences, USA). The sample was mixed and centrifuged to pellet insoluble material, then 1% Immobilized pH Gradient (IPG) buffer (pH 3–10 NL; GE Life Sciences) and 10 mM DTT were added to the supernatant before loading onto isoelectric focusing (IEF) strips (ReadyStrip, non-linear pH 3–10, 7 cm; Bio-Rad, USA) for 24 h passive rehydration. Proteins were electrophoresed in an Ettan IPGphor3 IEF system (GE Life Sciences, USA) under the following conditions: 100 V for 1 h, 300 V for 200 Volt-hours (V-h), 300 to 1000 V for 300 V-h, 1000 to 5000 V for 4000 V-h, and 5000 V for 1250 V-h. The IPG strip was then equilibrated for 10 min in reducing equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 1.5% DTT) followed by a second incubation for 20 min in alkylating equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 30% glycerol, 2% iodoacetamide). The IPG strip was then embedded on top of a 12.5% polyacrylamide gel (PowerPac Electrophoresis unit; Bio-Rad) and covered with 0.5% agarose. Second dimension electrophoresis was performed at 4°C for 1 h at 150 V per gel. The resulting gel was stained overnight with 0.2% colloidal Coomassie brilliant blue G250 (34% methanol, 3% phosphoric acid, 170 g/L ammonium sulfate, 1 g/L Coomassie blue G250), then destained in 1% acetic acid/H_2O. Visible spots were subsequently picked from the gel and digested overnight at 37°C using sequencing-grade trypsin (Sigma, USA). Briefly, gel spots were washed with ultrapure water, destained (40 mM NH_4CO_3/50% acetonitrile (ACN)) and dehydrated (100% ACN). Gel spots were rehydrated in 10 µL of 20 µg/ml proteomics-grade trypsin (Sigma-Aldrich) and incubated overnight at 37°C. Digests were eluted by washing the gel spots for 30 min with

each of the following solutions: 50 µL 50% ACN/1% formic acid (FA), followed by 50 µL 70% ACN/1% FA. The samples were then dried by evaporation using a vacuum centrifuge and reconstituted in 20 µL of 1% FA prior to analysis by LC-MS/MS.

3.4.3 HPLC

Reversed-phase (RP) HPLC analysis of crude venom was performed using a Shimadzu Prominence system. Venom (1 mg) was fractionated on a C_{18} column (4.6 × 250 mm, 5 µm particle size, 300 Å pore size) using a flow rate of 1 ml/min and the following gradient of solvent B (0.043% trifluoroacetic acid (TFA) in 90% ACN) in solvent A (0.043% TFA in water): 10% solvent B for 15 min, 10–65% solvent B over 50 min, 45–70% solvent B over 5 min. Fractions were collected manually, dried by evaporation using a vacuum centrifuge, then prepared for LC-MS/MS analysis as described below.

3.4.4 Protein identification using LC-MS/MS

To identify proteins present in the milked venom we used a bottom-up proteomics approach to analyze the digested 2DE gel spots, RP-HPLC fractions, as well as crude desalted venom. Reduction and alkylation of cysteine residues in venom proteins and peptides was performed as reported previously (180). Reduced/alkylated venom was incubated overnight at 37°C in 10 µL of 40 ng/µL proteomics-grade trypsin (Sigma) in 40 mM NH_4CO_3, pH 8. The digested reduced/alkylated samples were then resuspended in a final concentration of 1% FA and centrifuged for 15 min at 12,000 g prior to LC-MS/MS. For analysis of RP-HPLC fractions and in-gel digests, tryptic peptides were fractionated on an Agilent Zorbax stable-bond C_{18} column (2.1 mm × 100 mm, 1.8 µm particle size, 300 Å pore size) using a flow rate of 180 µl/min and a gradient of 1–40% solvent B (90% ACN, 0.1% FA) in 0.1% FA over 15 min on a Shimadzu Nexera UHPLC coupled with an AB SCIEX 5600 mass spectrometer equipped with a Turbo V ion source heated to 500°C. MS/MS spectra were acquired at a rate of 20 scans/s, with accumulation time of 0.25 ms, resulting in a cycle time of 2.3 s, and optimized for high resolution. Precursor ions with m/z of 300–1,800 m/z, a charge of +2 to +5, and an intensity of at least 120 counts/s were selected, with a unit mass precursor ion inclusion window of ± 0.7 Da, and excluding isotopes within ±2 Da for MS/MS. The crude venom digest was analyzed as above except using a gradient of 1–40% solvent B (90% ACN, 0.1% FA) in 0.1% FA over 60 min.

Mass spectra were searched against predicted coding sequences (CDSs) from the assembled transcriptome (see below) using ProteinPilot v4.5 (AB SCIEX). Searches were run as thorough identification searches, specifying tryptic digestion and the alkylation reagent as appropriate. Biological modifications and amino acid substitutions were allowed in order to maximize the identification of protein sequences from the transcriptome despite the inherent variability of toxins, potential isoform mismatch with the transcriptomic data, and to account for experimental artifacts leading to chemical modifications. We used a stringent detected protein threshold score of 1% false discovery rate (FDR) as calculated by decoy searches.

3.4.5 Transcriptome sequencing, assembly, and expression analysis

Tentacle tissue from *A. pulchella*, *A. tenebrosa*, *H. malu*, *M. doreensis* and *S. haddoni* were collected with tweezers and flash frozen before total RNA was extracted using TRIzol (Life Technologies) and enriched for mRNA using a DynaBeads Direct mRNA kit (Life Technologies) as described (257). mRNA (~ 350 ng) was supplied to the Institute for Molecular Bioscience Sequencing Facility for library preparation and sequencing. A paired-end cDNA library (180 bp insert size) was prepared using a TruSeq-3 library kit and sequenced on an Illumina NextSeq 500 (mid-output, 150 bp paired-end reads). In addition to the transcriptomes sequenced here, SRAs from *Aiptasia sp* (SRR1648361), *Anemonia viridis* (SRR1573633), *Anthopleura elegantissima* (SRR2300240), *Bolocera tuediae* (SRR504347), *Cnidopus japonicus* (Red form — SRR2134406; Yellow form — SRR2134407), *Hormathia digitata* (SRR504348), *Nematostella vectensis* (ERR1368849), and *Urticina eques* (SRR942796) were analysed. The resulting reads were trimmed using Trimmomatic v0.35 (183) to remove adapter sequences and low-quality reads. Window function-based quality trimming was performed using a windowsize of 75 and windowquality of 34, and sequences with a resulting length of <100 bp were removed. After quality control, paired-end sequences were *de novo* assembled into contigs using Trinity v2.0.6 (258) using default parameters. For inspection of assembled contigs, trimmed paired reads were mapped to the Trinity assembled reads using bowtie v2.2.6 (259) and visualised using the Integrative Genomics Viewer v2.4.x (260, 261).

3.4.6 Functional annotation of transcriptome

In order to identify potential toxin-like transcripts, we compared the translated transcriptome to all curated animal toxin sequences. Coding sequences (CDSs) were identified using the

Galaxy tool ‘Get open reading frames (ORFs) or coding sequences (CDSs)’ (189). A minimum CDS length cut-off of 30 residues was used to minimize the probability of not identifying short toxin CDSs. BLASTp searches of the resulting translated CDSs against the proteomics database we generated in addition to sea anemone toxins deposit to UniProt venom database (190) were performed with the upper-limit for the E-value set to 1E-6. Candidate toxin-like transcripts were further processed by removing redundant protein sequences using CD-HIT (191) as well as sequences not containing a signal peptide that could be detected using SignalP (version 4.1) (192). Finally, the toxin candidates were classified into categories according to predicted structure and/or function.

3.4.7 Functional annotation of proteomics data

CDSs from the venom transcriptome were used as a protein database for proteomic analyses. LC-MS/MS spectra were searched against the CDS database using ProteinPilot v4.5 (AB SCIEX). Sequences with less than two peptides with 95% confidence were excluded; the selected sequences from ProteinPilot were extracted using Galaxy tools. These sequences were then BLAST searched against the UniProt animal toxin database (www.uniprot.org/program/toxins) and annotated according to the methodology described for the transcriptome annotation. Sequences without matches to the UniProt toxin database were functionally annotated using a combination of BLAST searches against the NCBI non-redundant protein database and InterProScan (193) using Blast2GO (194). Sequences without known functional or structural motifs were considered "unknowns" (U). To identify the full molecular diversity contained within each of the putative toxin families, sequences identified in the milked venom were then BLAST searched against the translated transcriptome. Redundant protein sequences were removed using CD-HIT and only sequences containing a predicted signal peptide were considered as putative toxin candidates.

3.4.8 Phylogenetic Analyses

Translated nucleotide sequences of each toxin types were aligned using mafft v7.304b (262) and refined using the tool re-alignment. CLC Main Workbench v7.6.1 (CLCbio, QIAGEN, Denmark) were used for visualization. Maximum likelihood phylogenies reconstructed with IQ-Tree v1.5.5 (263) for each toxin type. Evolutionary models were estimated using ModelFinder (264), while support values were estimated by ultrafast bootstrap using 1000

iterations (265). Trees were displayed using the online tool ‘Interactive Tree of Life’ (iTOL) (266).

CHAPTER 4
PHAB toxins: a unique family of sea anemone toxins evolving via intra-gene concerted evolution defines a new peptide fold

Abstract

Sea anemone venoms have long been recognised as a rich source of peptides with diverse structure and pharmacological properties, but they still contain many uncharacterised bioactive compounds. Here we report the discovery, three-dimensional structure, activity, tissue localisation, and putative function of a novel peptide toxin that constitutes a new, sixth type of voltage-gated potassium channel (K_V) toxin from sea anemones. Comprised of just 17 residues, κ-actitoxin-Ate1a (Ate1a) is the shortest sea anemone toxin reported to date, and it adopts a novel three-dimensional structure that we named the Proline-Hinged Asymmetric β-hairpin (PHAB) fold. Mass spectrometry imaging and bioassays suggest that Ate1a serves a primarily predatory function by immobilising prey, and we show that this is achieved via inhibition of *Shaker*-type K_V channels. Ate1a is encoded as a multi-domain precursor protein that yields multiple identical mature peptides, which likely evolved by multiple domain duplication events in an actiniodean ancestor. Despite this ancient evolutionary history, the PHAB-encoding gene family exhibits remarkable sequence conservation in the mature peptide domains. We demonstrate that this conservation is likely due to intra-gene concerted evolution, which has not previously been reported for toxin genes. We propose that the concerted evolution of toxin domains provides a hitherto unrecognised way to circumvent the effects of the costly evolutionary arms race considered to drive toxin gene evolution by ensuring efficient secretion of ecologically important predatory toxins.

Significance

Sea anemone venoms are a rich source of peptide toxins that have found application as pharmacological tools and drug leads. Here we describe a new sea anemone toxin, Ate1a, which is the first representative of a previously undescribed peptide family that adopts a novel β-hairpin-like three-dimensional fold. Ate1a lacks the antimicrobial activity characteristic of many β-hairpin peptides, and instead it is used to paralyse envenomated prey. In contrast with most venom toxins, the Ate1a peptide family comprises a few, highly conserved genes encoding multiple copies of the mature toxin. These domains evolve by intra-gene concerted evolution, a process never before documented for animal toxins. We propose that this process represents an unrecognised strategy for efficient secretion of ecologically important toxins.

4.1 Introduction

Venoms are complex cocktails of bioactive molecules that disrupt the physiology of envenomated prey (267, 268). Although these toxins include a wide range of molecules, such as proteins, peptides, polyamines, and salts, the impressive molecular diversity of most invertebrate venoms is due to disulfide-rich peptides (269). In animals that rely on venom for prey capture, diet and foraging ecology are thought to be major drivers of toxin evolution, with the acquisition of resistance in prey countered by diversifying selection acting on toxin genes in the predator (270). As a result, the venoms of predatory animals tend to be highly diverse, often containing hundreds to thousands of unique bioactive toxins (269). One such group is sea anemones, which are benthic, sessile cnidarians that use venom for a variety of ecological functions, including prey capture, defence, digestion, and inter- and intraspecific competition.

Given the ecological importance of venom in sea anemones, and the fact that the cnidarian venom system has been evolving for >700 million years (227), it is not surprising that sea anemones have evolved a rich variety of venom toxins including enzymes, cytolysins, and neurotoxins (148, 235). Of these, disulfide-rich peptide neurotoxins constitute the largest molecular diversity. According to the classification system proposed by Mikov and Kozlov (271), at least 17 different peptide folds have been identified in sea anemone venoms (272), although recent proteomics studies suggest that they likely contain 30 or more (273).

In addition to being the most diverse components of sea anemone venoms, neuroactive peptides are also the most well studied. They have been used as tools for probing ion channel structure and function, and for developing novel therapies (235). For example, ShK, a venom peptide from the sea anemone *Stichodactyla helianthus*, recently completed Phase 1 clinical trials for treatment of autoimmune disease (274). Neurotoxins from sea anemone venoms act on a diverse range of ion channels, including acid-sensing ion channels (ASIC), transient receptor potential ion (TRP) channels, and voltage-gated sodium (Na_V) and potassium (K_V) channels. Of these, K_V toxins are the most diverse group, comprising 136 of the 320 annotated sea anemone toxins in UniProtKB. These K_V toxins are currently divided into five distinct types based on their sequence, disulfide-bridge pattern, and activity (94).

Here we describe the structure, activity, function and evolution of a new, sixth type of sea anemone K_V toxin. κ-Actitoxin-Ate1a (henceforth Ate1a), from venom of the Waratah sea anemone *Actinia tenebrosa,* is the shortest sea anemone toxin reported to date, and it adopts a

novel β-hairpin-like 3D fold. In contrast with many β-hairpin peptides, Ate1a lacks antimicrobial activity and instead serves a predatory function via potent inhibition of prey K_V channels. While most families of predatory toxins evolve via bursts of extensive duplication and diversification, this is not the case for the Ate1a toxin family, whose members remain remarkably well conserved despite their ancient evolution. Our data suggest that this extreme conservation is due to intra-gene concerted evolution, a process that has not been previously reported for any toxin family and which we propose is a hitherto unrecognised mechanism of maintaining efficient secretion of ecologically important toxins.

4.2 Results

4.2.1 Discovery of Ate1a

Specimens of *A. tenebrosa* were collected off the coast of North Stradbroke Island, Queensland, Australia (27°15 S, 153°15 E), and venom obtained by electrical stimulation (273). Fractionation of venom using reversed-phase chromatography revealed a conspicuous early-eluting peak containing an unusually low-mass component (Figure 4.1A), which we confirmed to be a disulfide-rich peptide by *de novo* sequencing using in-source dissociation (ISD) matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) (Figure 4.1B). The toxin, which we named Ate1a, is a 17-residue peptide with two disulfide bonds and an amidated C-terminus (RCKTCSKGRCRPKPNCG-NH_2), yielding a monoisotopic mass of 1887.93 Da. Ate1a is a novel peptide with no BLAST hits in UniProtKB or NCBI databases.

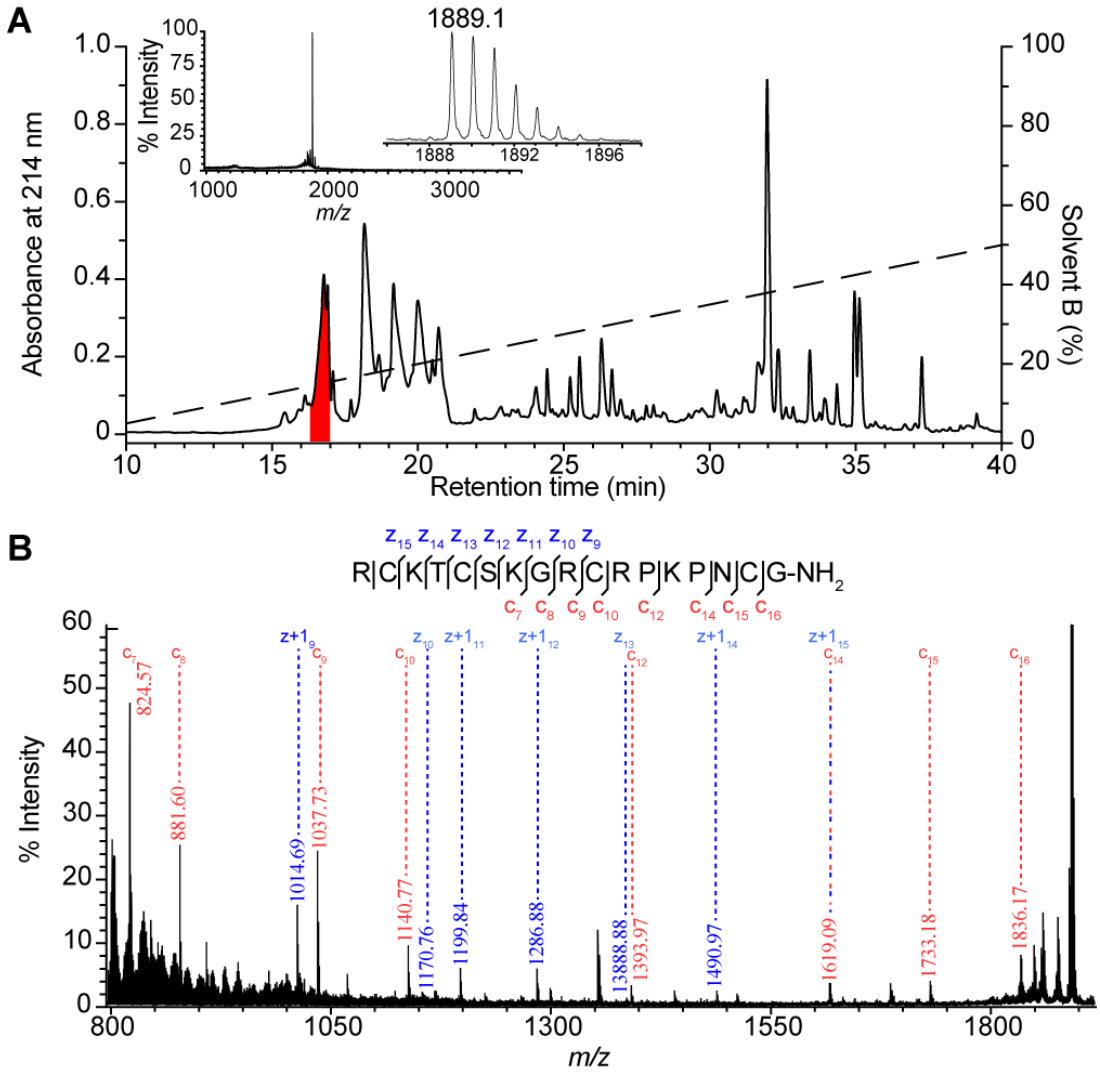


Figure 4.1 – Isolation and sequencing of Ate1a. (A) C18 RP-HPLC chromatogram showing fractionation of crude A. tenebrosa venom. The early-eluting peak containing Ate1a is highlighted in red. Inset shows average mass and isotope family for Ate1a. **(B)** De novo sequencing of Ate1a using ISD-MALDI MS.

To confirm the amino acid sequence of Ate1a and identify any venom homologues, we constructed a transcriptome from tentacles actively regenerating venom, as described previously (273). *De novo* assembly with Trinity yielded 87,485 contigs, which translated to 457,470 potential coding sequences (CDS). A BLAST search was used to identify the transcript encoding Ate1a, and this returned a single contig containing a partial CDS with multiple copies of a peptide domain encoding a sequence identical to that determined by ISD-MALDI-MS. Analysis of remapped reads revealed that this contig represents two unique transcripts whose CDS differ by two synonymous and two non-synonymous mutations in the propeptide regions of the Ate1a preproprotein (Appendix Figure V.i, NCBI SRA accession SRR6282389).

4.2.2 Ate1a belongs to a novel peptide family that evolves by intra-gene concerted evolution

We used the complete sequence of the Ate1a precursor to search for related sequences. A BLAST search against the UniProtKB and GenBank nr databases returned no significant hits, but BLAST searches against the NCBI expressed sequence tag (EST) database returned one full-length and two partial Ate1a-like prepropeptide sequence in *Anemonia viridis*. Similarly, a BLAST search against our published tentacle transcriptome from *Stichodactyla haddoni* (273) yielded three unique contigs, including one full-length prepropeptide. These species include two separate families (Actinidae and Stichodactylidae), suggesting that the Ate1a toxin-gene family arose in a common ancestor of the superfamily Actinioidea (2).

All identified Ate1a-like prepropeptides are comprised of the same set of domains separated by dibasic cleavage sites: a signal peptide, one or two cysteine-containing propeptide domains, and three cysteine-free propeptide domains that each precede an Ate1a-like domain (Figure 4.2A). The domain architecture is also the same for all prepropeptides except for the position of the cysteine-containing propeptide, which in Stichodactylidae (*S. haddoni*) is found as a single copy immediately following the signal peptide, but in Actinidae (*A. tenebrosa* and *A. viridis*) exists as two copies that each follow the first two Ate1a-like domains. This suggests that the Ate1a precursor gene underwent early domain duplication events followed by either multiple domain deletions or convergent deletions and duplications. Strikingly, however, the domains share almost 100% nucleotide identity with corresponding domains within the same prepropeptide (Figure 4.2B). Phylogenetic analysis revealed that domains from each species form well-supported clades with respect to those from other species (Figure 4.2C). This suggests that the extreme sequence conservation observed in domains of the Ate1a gene family is due to concerted evolution, an evolutionary phenomenon that, to our knowledge, has previously only been demonstrated for a single animal toxin gene family (275), but never for within-gene toxin domains.

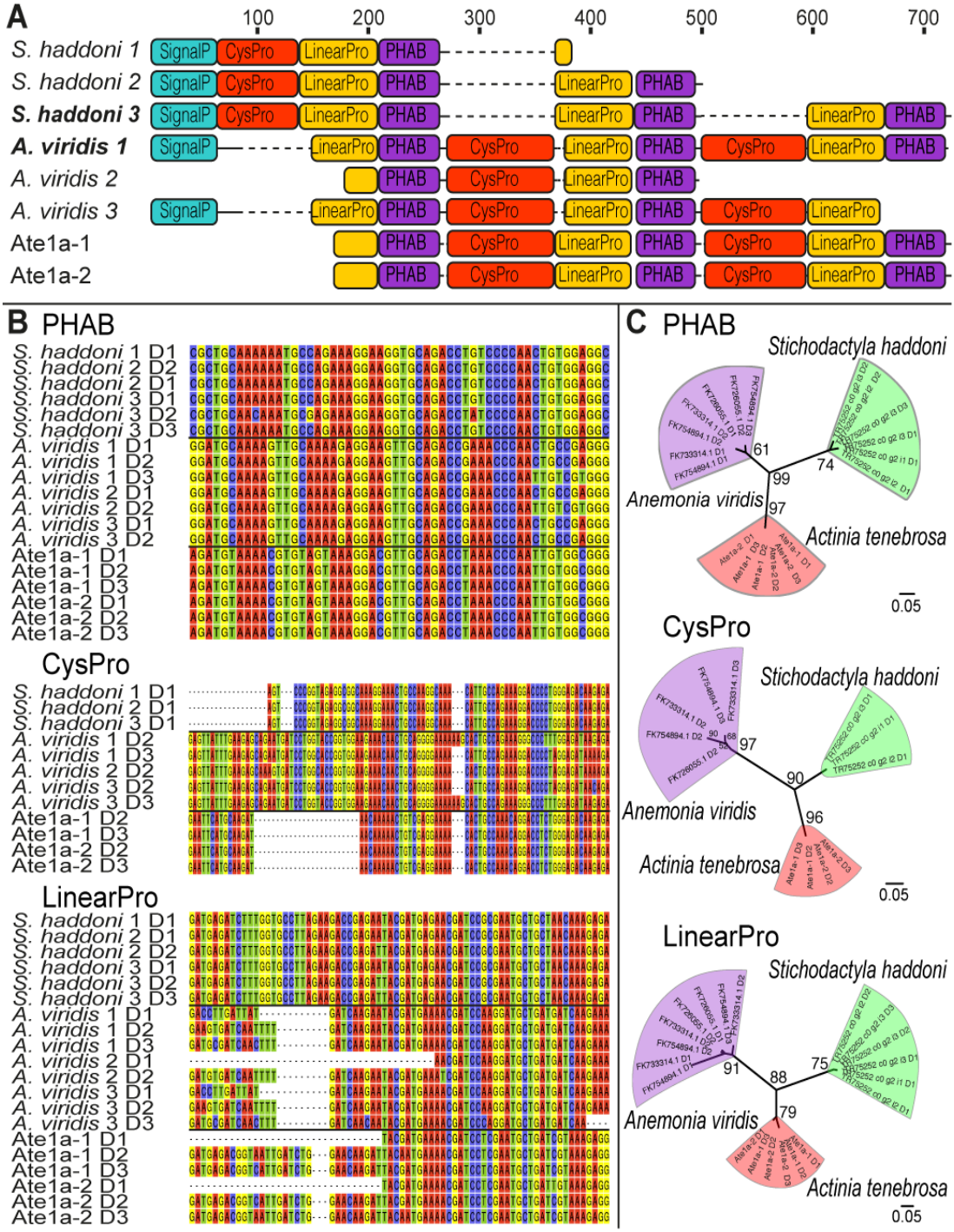


Figure 4.2 – Domain architecture and evolution of Ate1a precursors. (A) Domain architecture of Ate1a and Ate1a-like contigs. Prepropeptides are composed of a signal peptide (SignalP), one or two cysteine-containing propeptide domains (CysProP), and three cysteine-free propeptide domains (LinearProP) that each precede an Ate1a-like PHAB domain. **(B)** Nucleotide sequence alignments for each domain. **(C)** Maximum likelihood phylogenetic reconstructions for each domain. Bootstrap support values are shown at the nodes, while horizontal bars indicate genetic distance. Sequence accessions are for *S. haddoni* (1–3) TR75252_c0_g2_i1–3 and *A. viridis* (1) FK754894, (2) FK726055, and (3) FK733314.

4.2.3 Ate1a defines a new peptide fold

The unique primary structure of Ate1a prompted us to characterise its solution structure using NMR spectroscopy. We synthesised Ate1a using solid phase peptide synthesis and confirmed correct folding of the synthetic product by HPLC co-elution with native peptide (Appendix Figure VI.i).

The 3D structure of Ate1a (Figure 4.3) was determined using homonuclear NMR methods, and statistics for the ensemble of structures are shown in Appendix Table XI.i. MolProbity analysis (276) revealed that the structure has excellent stereochemical quality, with no steric clashes and ~90% of residues in the most favoured Ramachandran region. The precision of the structure, however, is not very high (backbone RMSD 1.11± 0.28 Å), suggesting that it is highly dynamic, particularly within the longer loop 3. This may be due the presence of two proline residues, which leads Ate1a to adopt two distinct conformations (Figure 4.3A). Both conformations adopt a fold similar to that of β-hairpin-like peptides (277), where the C- and N- termini are connected via two semi-parallel disulfide bonds (C1-C4 and C2-C3). One face of the toxin has a high proportion of positively charged residues (Figure 4.3B) whereas the opposite face is rich in hydrophobic residues (Figure 4.3C).

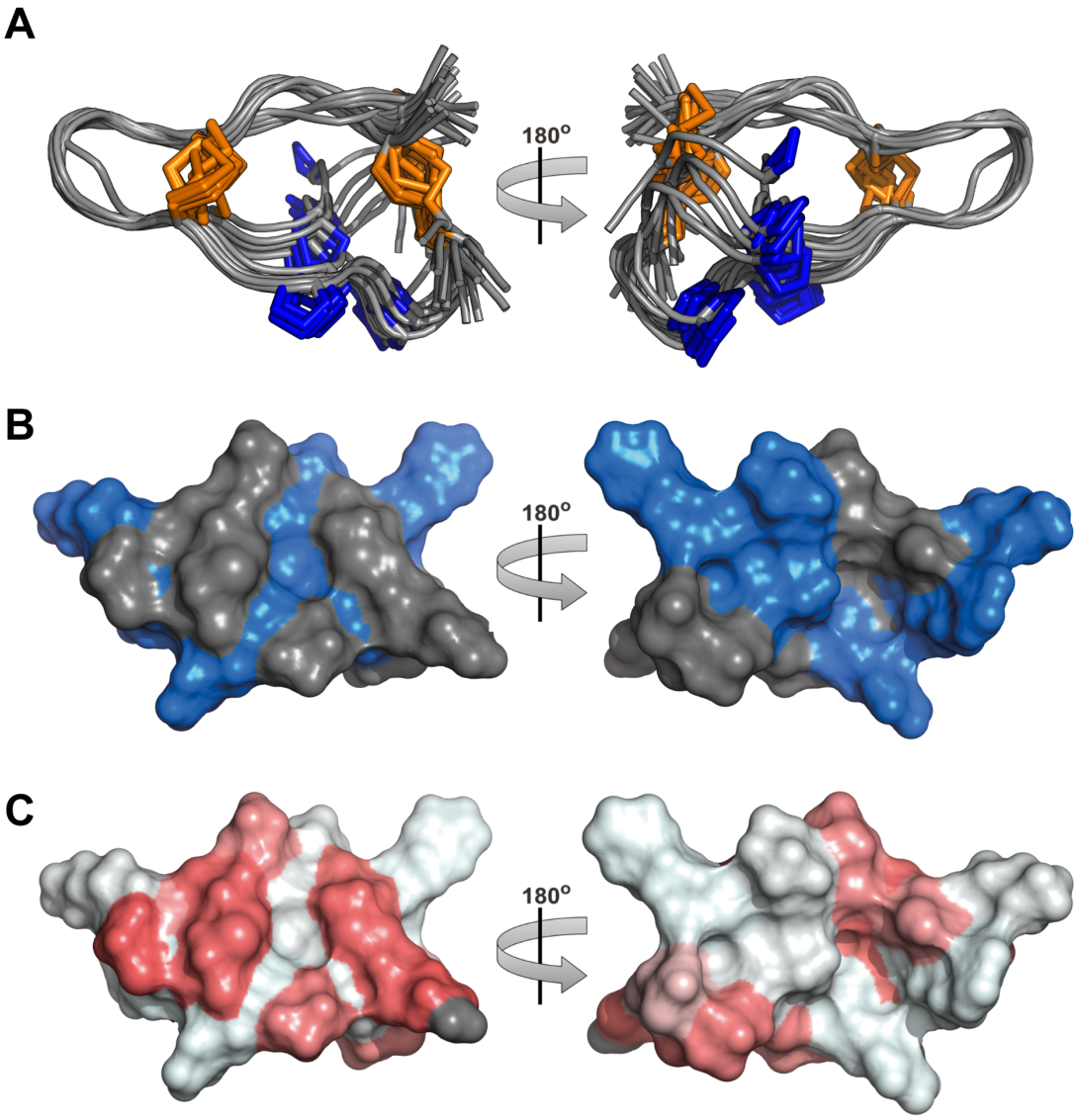


Figure 4.3 – 3D structure of Ate1a. (A) Solution structure of Ate1a (ensemble of 20 structures; PDB code 6AZA). Disulfide bonds are highlighted in orange and proline side chains are shown in blue. **(B)** Surface representation of Ate1a with cationic and uncharged residues shown in blue and grey, respectively. **(C)** Surface representation of Ate1a showing relative hydrophobicity, which increases from white to red.

Although Ate1a displays a hairpin-like structure (Figure 4.4A), it is neither a true hairpin scaffold nor similar to any other previously described hairpin-like peptide fold. The two disulfide-enclosed loops of Ate1a differ substantially in length, with loop 1 containing just two residues compared to five in loop 3. In combination with the two prolines in loop 3, this asymmetry prevents the formation of secondary structures characteristic of other disulfide-enclosed hairpin-like structures such as β-hairpin antimicrobial peptides (AMPs) (277) (Figure 4.4B) or the cystine-stabilised α/α (CSαα) fold (278) (Figure 4.4C). Ate1a also differs from

other hairpin-like folds found in animal toxins, such as the boundless β-hairpin (BBH) (24) (Figure 4.4D) and disulfide-directed hairpin (DDH) fold (279) (Figure 4.4E), or the two-disulfide fold of RhTx from venom of the centipede *Scolopendra subspinipes* (280) (Figure 4.4F). Thus, Ate1a is the prototypic member of a previously undescribed peptide fold that we coined the ***P**roline-**h**inged **a**symmetric **β**-hairpin-like* (PHAB) fold.

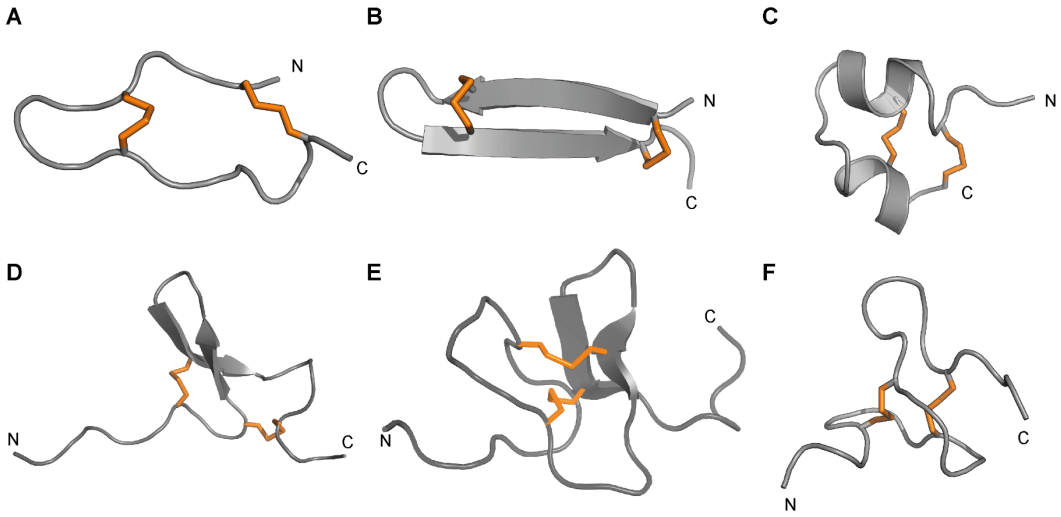


Figure 4.4 – Ate1a is the first member of the new PHAB fold. Comparison of the PHAB fold with other peptide folds containing two disulfide bonds and a similar number of residues (16-29 residues). Disulfide bonds are shown as orange tubes and N- and C-termini are labelled. (A) Ate1a; **(B)** β-hairpin fold represented by the spider peptide gomesin (PDB 1KFP); **(C)** CS α/α motif represented by the scorpion toxin κ-hefutoxin1 (PDB 1HP9); **(D)** Boundless β-hairpin motif represented by sea anemone toxin π-AnmTX Ugr 9a-1 (PDB: 2LZO); **(E)** Disulfide-directed hairpin represented by scorpion toxin U_1-Liotoxin-Lw1a (PDB 2KYJ); **(F)** Unstructured two-disulfide peptide fold represented by centipede toxin RhTx (PDB 2MVA).

4.2.4 Ate1a represents a new type of sea anemone K_V-toxin

Many Arg/Lys-rich, disulfide-stabilised β-hairpin peptides (e.g., gomesin and tachyplesin-1) function as AMPs in the innate immune system. They often have high affinity for lipid membranes and possess both anticancer and antimicrobial activity (281). Although Ate1a does not adopt a typical β-hairpin fold, it is highly positively-charged. However, Ate1a had no antimicrobial activity at concentrations up to 256 μg/mL (**Error! Reference source not found.**Appendix Table XII.i). Similarly, Ate1a was not cytotoxic or cytolytic against cultured

human cancer cell lines or erythrocytes (Appendix Figure VII.i). Consistent with these results, Ate1a displayed only weak affinity for, and rapid dissociation from, lipid membranes compared to gomesin (282) (Appendix Figure VIII.i). Taken together, Ate1a's lack of antimicrobial and cytolytic activity, as well as its low affinity for lipid membranes, suggests that it does not play a role in defence against pathogens.

Ion channels are the most common molecular target of disulfide-rich venom peptides, and we therefore used electrophysiology to screen Ate1a against eight Na_V channels, twelve K_V channels, and four ASIC subtypes. Ate1a was found to selectively target several members of the *Shaker* subfamily of K_V channels; at 3 μM it inhibited currents mediated by $K_V1.1$ (84% ± 4%), $K_V1.2$ (94% ± 3%), $K_V1.3$ (38% ± 4%), $K_V1.6$ (92% ± 2%) and Shaker IR (23% ± 2%) channels (Appendix Figure IX.iA). No activity was observed on other channels at the same concentration (Appendix Figure X.i). Fitting of the Hill equation to concentration–response curves for $K_V1.1$, $K_V1.2$, $K_V1.3$ and $K_V1.6$ yielded IC_{50} values of 353 nM, 146 nM, 3051 nM and 191 nM, respectively (Appendix Figure IX.iB). Thus, given its unique sequence and structure, Ate1a represents a new, sixth type of sea anemone K_V toxin.

4.2.5 Ate1a is a toxin with a predatory function

Although the pharmacological activity of a toxin can provide clues to its ecological function, it is not by itself definitive. However, the near-universal distribution of nematocytes in the epithelium of sea anemones means that toxin function can be inferred from tissue distribution (170, 283, 284). We therefore investigated the tissue distribution of Ate1a using MALDI-MS imaging (MSI), which allows visualization of the spatial distribution of unlabelled low mass biomolecules (1–20 kDa) (285-287). A peak corresponding to the average mass of Ate1a was clearly observed in MALDI-TOF-MSI spectra acquired from cross-sectioned *A. tenebrosa*. The identity of this peak was further supported by on-tissue gas-phase reduction and alkylation, which resulted in a peak shift matching the alkylation of four cystines (Figure 4.5A). Finally, ultra-high mass resolution analysis by MALDI-FT-ICR-MSI allowed us to fit the predicted isotope structure of Ate1a to the observed spectra and confirm its identity (Figure 4.5B). MSI revealed that Ate1a is non-uniformly distributed within the body of *A. tenebrosa*, with almost exclusive localization in tentacles (Figure 4.5C), suggesting that it is involved in prey capture. Ate1a mass signals were weak or absent in actinopharynx, mesenterial filaments, and

gastrovascular cavity, indicating that Ate1a does not play a role in prey digestion. Moreover, *A. tenebrosa* normally retracts its tentacles in response to disturbances, and thus the weak Ate1a signal in the trunk region indicates it is not primarily involved in defence.

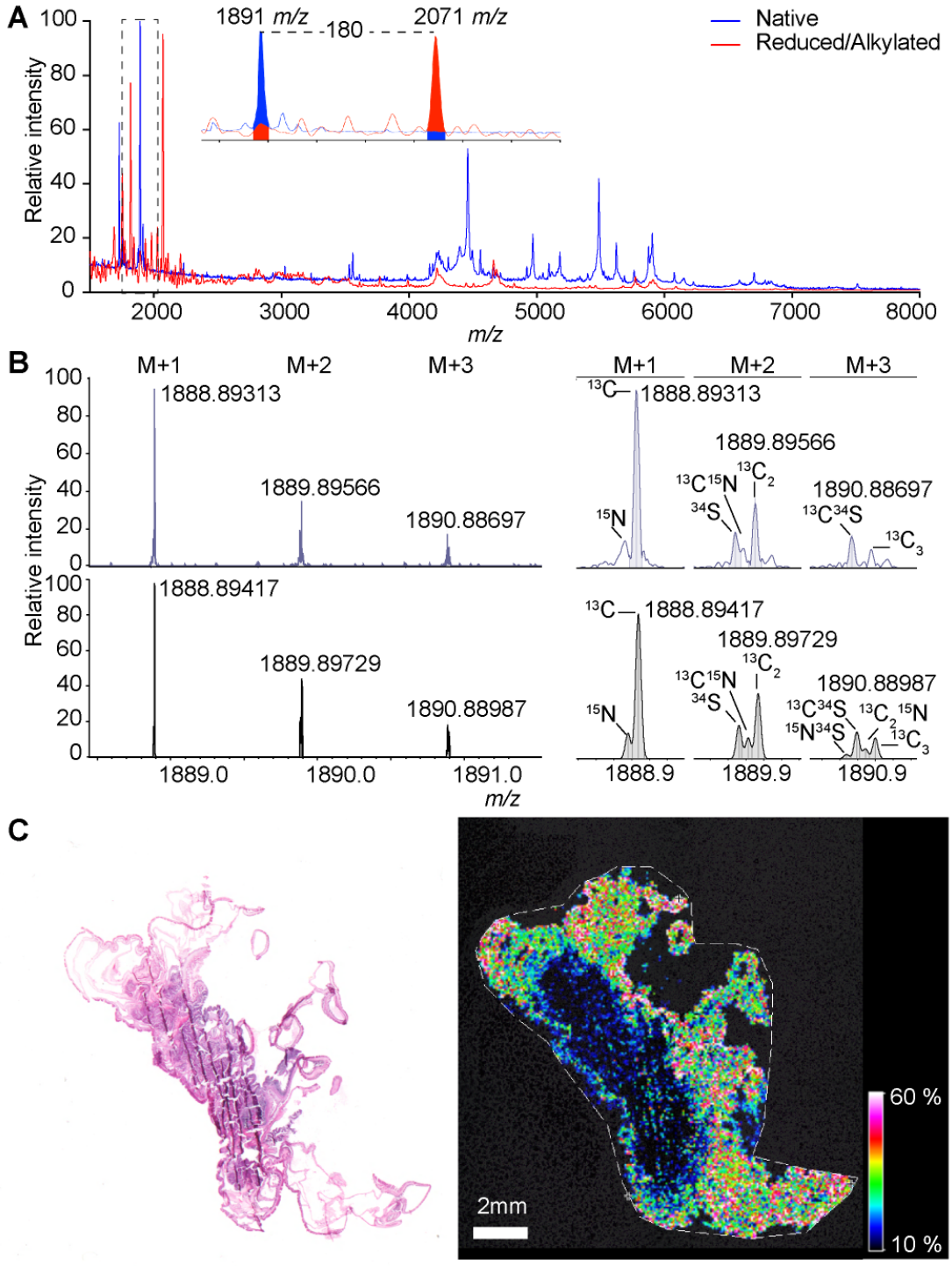


Figure 4.5 – Tissue distribution of Ate1a determined using MSI. (**A**) MSI linear positive mode spectra acquired from a cross-sectioned animal, with peaks corresponding to Ate1a filled in. The spectrum of native tissue is blue while the spectrum obtained after on-tissue gas-phase reduction and alkylation is shown in red. Inset shows a mass difference of 180 Da, corresponding to ethanolylation of four cysteine residues. (**B**) Ultra-high mass resolution analysis of the peak corresponding to Ate1a acquired by MALDI-FT-ICR-MSI at a resolution of 16,000,000 and resolving power at 1890 *m/z* of > 500,000, showing observed (top) and calculated (bottom) spectra. (**C**) Left: Histological image of the sea anemone section used for MSI experiments, stained with hematoxylin and eosin. Right: distribution of the peak corresponding to the average mass of Ate1a as observed by MALDI-TOF MSI.

In addition to nematocytes, sea anemones also produce toxins in ectodermal gland cells (10). Unlike nematocytes, which are stinging cells that inject venom, gland-cell toxins are released into the water and absorbed by prey. To determine which cell type produces Ate1a, we conducted toxicity bioassays using brine shrimp and amphipods, the latter being a major prey item of *Actinia* spp. (288, 289). Injection of Ate1a into amphipods resulted in impaired swimming followed by contractile paralysis. In contrast, Ate1a did not affect either species when dissolved into the medium (artificial sea water). Taken together, our data suggest that Ate1a is a neurotoxin produced in nematocytes and used primarily for prey capture.

4.3 Discussion

4.3.1 Ate1a is the founding member of a new, sixth type of sea anemone K_V toxin

Although sea anemone venoms are a rich source of bioactive peptides, recent omics studies have highlighted how little we still know about their composition, function, and evolution (170, 273). Here we described the discovery and functional characterization of a new peptide class from venom of the sea anemone *A. tenebrosa*, one of the most commonly encountered sea anemones in intertidal zones around Australia and New Zealand (290). Ate1a has a primary structure unlike any previously described peptide, and assumes a unique 3D fold that is reminiscent of β-hairpin AMPs (277).

In contrast to β-hairpin AMPs, the 3D structure of Ate1a is devoid of regular secondary structure. Instead, the asymmetry of the two sides of the β-hairpin-like structure of Ate1a prevents β-sheet formation, and distinguishes the 3D structure from previously described two-disulfide peptide folds (Figure 4.4). The longer of the two 'loops' is also highly dynamic (Figure 4.3A), a property facilitated by the presence of two prolines that are conserved in all identified Ate1a homologues. Proline-containing peptides have the ability to populate two discrete conformations, and this *cis-trans* conformational switch works like a hinge that can potentially serve as a precise regulator of biological function (291, 292). While proline hinges play a diversity of roles in protein biology, one of these roles is reorienting surface loops to modulate protein binding surfaces and in turn ligand recognition (293). Thus, we predict that the proline-hinged loop of Ate1a represents a region that is important for the function of this toxin family. Given the structural and likely functional importance of this structural feature, we named this new structural scaffold the "*proline-hinged asymmetric β-hairpin-like*" (PHAB)

fold.

Reflecting the structural distinctiveness of the PHAB fold from β-hairpin-like peptides, Ate1a does not have antimicrobial, antifungal, or cytolytic activity, nor strong affinity for lipid membranes (Appendix Figure VII.iAppendix Figure VIII.i). Instead, it is a potent inhibitor of *Shaker*-type K_V channels, with nanomolar potency on $K_V1.1$, $K_V1.2$ and $K_V1.6$. K_V channels play crucial roles in neuronal signalling, muscle contraction and secretion (294), and hence they are a common target of animal toxins. Many venomous taxa have convergently evolved toxins that target K_V channels to induce paralysis, general hyperexcitability, cardiac disorders, convulsions and death (267). This is also the case in sea anemones, where K_V toxins are represented by five unique peptide folds: ShK (type 1), Kunitz-domain (type 2), β-defensin-like (type 3), boundless β-hairpin (type 4), and an unknown fold predicted to form an inhibitor cystine knot (type 5) (94). The PHAB fold is unlike any of these structural scaffolds, and therefore it represents a new, sixth type of sea anemone K_V toxin.

4.3.2 Ate1a is a predatory toxin produced in nematocytes

Although Ate1a is a novel K_V toxin, correlating toxin activity with ecological function is often not straightforward (295). However, like other cnidarians, sea anemones lack a centralised venom delivery system, and instead rely on localised production of toxins to complement their functional anatomy (170, 283, 284). In *A. tenebrosa*, toxins are produced in five tissues and regions that have distinct ecological functions: acrorhagi (aggressive intraspecific encounters), tentacles (prey capture and immobilisation), mesenteric filaments (used principally in digestion), column (external defence after retracting tentacles), and actinopharynx (prey immobilisation and digestion). In addition, sea anemones produce toxins in two distinct cell types that deliver venom by either injection (nematocytes) or absorption following secretion into the water column (gland cells) (10). Ate1a is found predominantly in the tentacles of *A. tenebrosa* (Figure 4.5), which is suggestive of a predatory function. Moreover, Ate1a impaired swimming and led to paralysis and death when injected in amphipods, a major prey of *Actinia* species, but had no effect when dissolved into the medium. We conclude that Ate1a is a predatory toxin that cannot reach its K_V targets without being inoculated into prey by nematocysts.

4.3.3 PHAB toxins evolve via intra-gene concerted evolution

Venom proteins are thought to evolve via toxin recruitment events, whereby a gene encoding a normal body protein is duplicated and expressed in the venom-producing tissue (270). Functionally important toxin types are reinforced through duplication and diversification, and this is considered a hallmark of toxin evolution in predatory venoms, where toxins evolve continuously to counter acquisition of prey resistance (270). Although recent research suggests that venoms evolve via a two-step process in which initial rapid toxin diversification is followed by periods of purifying selection due to the metabolic costs of diversifying selection (220, 296), predatory toxins nevertheless tend to be part of large, highly diverse gene families. Strikingly, however, this diversity is entirely absent in the PHAB gene family, despite their likely role in predation. Instead, its members are highly conserved and consist of just 2–3 almost identical copies in each species (Figure 4.2B).

The sequence conservation at the nucleotide level is not limited to between-gene copies of each species, but extends to the domains encoded by each transcript. Despite the emergence and domain duplication of the PHAB fold in an actinioidean ancestor, all four domain types (signal peptide, two propeptide domains, and PHAB domain; Figure 4.2A) are remarkably well conserved. Furthermore, the nucleotide sequences encoding each domain type are more similar to the respective domains contained on the same transcript than to corresponding domain copies in other species (Figure 4.2C). This form of domain conservation is likely to have occurred by concerted evolution, an evolutionary process driven by continuous recombination that results in homogenisation of genetic variance across gene copies and so-called 'horizontal evolution' (297). Although this process has been described for a number of gene families, including Na_V Type I toxins from *Nematostella vectensis* and *Actinia equina* (275), it is considered rare for intra-gene protein domain repeats (298), and has never previously been reported within toxin gene domains.

In contrast with the general view of gene duplication as a facilitator of toxin gene diversification, recent studies have suggested that gene duplication may be of immediate importance for increased expression levels rather than generation of sequence diversity (299). Similarly, the concerted evolution of Na_V Type I toxins from *N. vectensis* and *A. equina* has been suggested to confer a selective advantage through a 'dosage' effect of gene expression (275). In the PHAB gene family, this lack of high gene-copy numbers is compensated for by encoding multiple, identical toxin precursors, thereby effectively multiplying toxin expression

levels. Concerted evolution may also facilitate 'transmission' of advantageous mutations from a single toxin gene locus to other loci, or preventing the loss of highly effective toxins. Reflecting this, concerted evolution of protein domain repeats has been proposed to be triggered by arms-race type co-evolution (298).

Thus, concerted evolution of toxin-domain repeats may provide a hitherto unrecognised mechanism of circumventing the effects of the metabolically expensive evolutionary arms race typically considered to drive toxin gene evolution. In the case of Ate1a and other members of the PHAB family, this has led to efficient secretion of structurally unusual but ecologically important, predatory K_V toxins.

4.4 Materials and Methods

4.4.1 Venom Collection and Fractionation

Sea anemones were collected at North Stradbroke Island, Queensland, Australia (27°15 S, 153°15 E) then housed in aquaria at The University of Queensland. Venom was obtained after a starvation period of at least 48 h. Briefly, the sea anemone was rinsed, placed in a minimal volume of artificial seawater, and nematocysts induced to discharge by electrical stimulation (179). The water, which contains the venom was lyophilized and then the venom was fractionated on a C_{18} column (4.6 × 250 mm, 5 μm particle size, 300 Å pore size) using a flow rate of 1 mL/min and the following gradient of solvent B (0.043% trifluoroacetic acid (TFA) in 90% acetonitrile (ACN)) in solvent A (0.043% TFA in water): 10% solvent B for 15 min, 10–65% solvent B over 50 min, 45–70% solvent B over 5 min.

4.4.2 Mass Spectrometry

4.4.2.1 MALDI-TOF/TOF Mass spectrometry

Peptide masses in lyophilized RP-HPLC fractions were analysed by MALDI-TOF MS using an AB SCIEX 5800 MALDI-TOF/TOF mass spectrometer. RP-HPLC fractions were mixed 1:1 (v/v) with α-cyano-4-hyroxy-cinnamic acid (CHCA) (7.5 mg/mL in 50/50 ACN/H_2O, 0.1% TFA). MALDI–TOF mass spectra were collected in reflector positive mode and reported masses are monoisotopic $M+H^+$ ions. For sequencing of intact Ate1a by in-source dissociation TOF-MS, 1,5-diaminonaphthalene (1,5-DAN) was used as a reductive matrix (300). The

sample was mixed 1:1 (v/v) with 1,5-DAN (15 mg/mL in 50/50 ACN/H_2O, 0.1% formic acid (FA)), and spectra were interpreted manually.

4.4.2.2 MALDI Mass Spectrometry Imaging

MALDI-MSI was guided by published protocols (301, 302) but with sample preparation optimized as recently described (303). Briefly, specimens of *A. tenebrosa* were left in 50% RCL2/ethanol at room temperature overnight, sequentially dehydrated into ethanol (3 × 15 min at each concentration), cleared in xylene for 30 min, and embedded in paraffin wax. A whole embedded animal was sectioned transversally at 7 µm thickness. Sections were de-paraffinized by careful washing with xylene, and optically imaged prior to applying CHCA (7 mg/mL in 50% ACN, 0.2% TFA) using a Bruker ImagePrep automated matrix sprayer. The on-tissue reduction and alkylation of cystines was carried out on de-paraffinized tissue sections using a volatile reaction protocol described previously (304), but with a 3.5 mL reaction volume in a 50 mL Corning Falcon tube (Thermo Fisher).

FlexControl 3.3 (Bruker) was used to operate an UltraFlex III TOF-TOF mass spectrometer (Bruker) in linear positive mode, with the range set to 1,000–20,000 m/z. A small laser size was chosen to achieve a spatial resolution of 50 µm, and matrix ion suppression was enabled up to 980 m/z. Individual MSI experiments were performed using FlexImaging 4.0 (Bruker). FlexImaging was used to establish the geometry and location of the section on the slide based upon the optical image, choose the spatial resolution, and call upon FlexControl to acquire individual spectra, accumulating 200 shots per raster point. FlexImaging was subsequently used to visualize the data in 2D ion-intensity maps, producing an averaged spectrum based upon the normalized individual spectra collected during the experiment.

For ultra-high mass resolution MSI we used a SolariX XR 7T FT-ICR mass spectrometer Bruker-Daltonik, Bremen, Germany) was used and operated in the positive ion mode. Data size was set to 1M across the mass range 400–6000 m/z. MALDI source was set to a laser power = 50%, a total of 500 shots per scan at a frequency of 2kHz, smart walk was enabled with a width of 90 µm. The Collision Cell RF Frequency was set to 1.4 MHz, Collision RF Amplitude 1100 Vpp, Transfer Optics Time of Flight = 1.5 ms at a frequency of 2 MHz with RF Amplitude = 400 Vpp. The sweep excitation was set to 20%. For isotopic fine structure analysis, data size was set to 4M across the mass range 200–3000 m/z. Data was collected and averaged across 8 scans. MALDI source was set to a laser power = 50%, a total of 5000 shots per scan at

frequency of 2 kHz, the laser spots was manually moved across sample area. For isolation the quadrupole was set to 1890.00 with an isolation window of 5 m/z. The Collision Cell RF Frequency was set to 2 MHz and Collision RF Amplitude set to 1200 VPP. The Transfer Optics were set to a Time of flight = 1.5 ms, Frequency set to 4 Mhz with RF Amplitude set to 400 Vpp. Sweep excitation set to 19%. For data analysis Bruker (Bruker-Daltonik, Bremen, Germany) DataAnalysis 5.0 and for image analysis Bruker FlexImaging 5.0 and Bruker SCiLS Lab 2017a were used.

4.4.3 Transcriptomics

Tentacle tissue from *A. tenebrosa* was collected with tweezers and flash frozen before total RNA was extracted using TRIzol (Life Technologies) and enriched for mRNA using a DynaBeads Direct mRNA kit (Life Technologies) as described (257). mRNA (350 ng) was supplied to the Institute for Molecular Bioscience Sequencing Facility for library preparation and sequencing. A paired-end cDNA library (180 bp insert size) was prepared using a TruSeq-3 library kit and sequenced on an Illumina NextSeq 500 (mid-output, 150 bp paired-end reads). The resulting reads were trimmed using Trimmomatic v0.35 (183) to remove adapter sequences and low-quality reads. Window function-based quality trimming was performed using a windowsize of 75 and windowquality of 34, and sequences with a resulting length of <100 bp were removed. After quality control, paired-end sequences were *de novo* assembled into contigs using Trinity v2.0.6 (258) using default parameters. For inspection of assembled contigs, trimmed paired reads were mapped to the Trinity assembled reads using bowtie v2.2.6 (259) and visualised using the Integrative Genomics Viewer v2.4.x (260, 261). Raw sequences reads (SRA: SRR6282389) and Trinity assembled contigs have been deposited with links to BioProject accession number PRJNA414357 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). Coding sequences (CDSs) were identified using the Galaxy tool 'Get open reading frames (ORFs) or coding sequences (CDSs)' (305). A minimum CDS length cut-off of 30 residues was used to minimize the probability of not identifying short toxin CDSs. The sequence determined by ISD-MALDI-MS was used to search the translated transcriptome using NCBI BLAST+ blastp (306).

4.4.4 Evolution of Ate1a

The Ate1a prepropeptide sequence was used to search for homologues in UniProtKB, NCBI nr and EST, and a tentacle transcriptome of *S. haddoni* (273) using NCBI BLAST+ blastp. Nucleotide sequences were retrieved and aligned using mafft v7.304b (262); domains were extracted using CLC Main Workbench v7.6.1 and maximum likelihood phylogenies reconstructed with IQ-Tree v1.5.5 (307) for each domain type. The evolutionary model (FLU+G4) was determined using ModelFinder (308), and support values estimated by ultrafast bootstrap using 10000 iterations (309).

4.4.5 Ate1a Synthesis

Fmoc amino acids and O-(1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) were from Iris Biotech (Marktredwitz, Germany), dimethylformamide (DMF) and diisopropylethylamine (DIEA) were from Auspep (Melbourne, Australia). All other reagents were obtained from Sigma Aldrich.

Ate1a (H-RCKTCSKGRCRPKPNCG-NH_2) was assembled on a 0.1 mmol scale using Fmoc chemistry on a Symphony automated peptide synthesizer (Protein Technologies Inc., Tucson, AZ, USA) on Fmoc-Rink-amide polystyrene resin (substitution value 0.67 mmol/g). Fmoc deprotections were accomplished by treatment with 30% piperidine/DMF (1 × 1 min, then 1 × 3 min). Amino acids were coupled using five equivalents of Fmoc amino acid/HCTU/DIEA (1:1:1) relative to resin loading (1 × 4 min then 1 × 8 min). Side-chain protecting groups used were: Arg(Pbf), Asn(Trt), Cys(Trt), Lys(Boc), Ser(tBu) and Thr(tBu). Cleavage from the resin and removal of side-chain protecting groups was achieved with 95% TFA/2.5% TIPS/ 2.5% H_2O for 2 h at room temperature. TFA was removed under a stream of nitrogen, then crude product was precipitated using cold diethyl ether (Et_2O). The product was washed with Et_2O, redissolved in 50% ACN/0.1% TFA/H_2O and lyophilised. ESI-MS (*m/z*): calc. (avg) 632.1 $[M+3H]^{3+}$, found 632.0. The crude product was oxidised by stirring in 0.1 M NH_4HCO_3 (pH 8.1) at room temperature for 3 days to give a single major isomer that was isolated by preparative HPLC. ESI-MS (*m/z*): calc. (avg) 630.8 $[M+3H]^{3+}$, found 631.0.

RP-HPLC solvent A was 0.05% TFA/H_2O and solvent B was 0.043%TFA/ 90% ACN/H_2O. Analytical HPLC was performed on a Shimadzu LC20AT system using a Thermo Hypersil GOLD C18 2.1 × 100 mm column and a gradient of 0–30% B over 30 min at flow rate of 0.3

mL/min. Absorbance was recorded at 214 nm. Preparative HPLC was performed on a Waters 600 system using a Vydac 218TP 22 × 250 mm column with a gradient of 0–30% B over 30 min at a flow rate of 16 mL/min. Mass spectra were recorded in positive ionisation mode on an AB SCIEX API 2000 triple quadrupole mass spectrometer.

4.4.6 Ate1a Structure

The solution structure of Ate1a was determined using 2D NMR spectroscopy. The NMR sample contained 1 mM synthetic Ate1a in 20 mM sodium phosphate, pH 6 and 5% D_2O. Spectra were acquired at 10°C on a Bruker AVANCE 600 MHz spectrometer equipped with a cryoprobe (Bruker, Billerica MA, USA). Resonance assignments were accomplished using a combination of 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C HSQC, 2D ^{1}H-^{1}H TOCSY and 2D ^{1}H-^{1}H NOESY spectra. The 2D NOESY spectrum (mixing time of 350 ms) was used to obtain interproton distance restraints. Spectra were processed using Topspin 3.5.b.91 pl 7 and analysed using CcpNmr Analysis 2.4.1 (310). 97.8% of resonances were assigned and they have been deposited in the BioMagResBank (accession number 30342). Dihedral-angle restraints were derived using TALOS-N (310) and the restraint range set to twice the estimated standard deviation. The NOESY spectrum was manually peak-picked, then the peak list was automatically assigned and structures calculated using the torsion angle dynamics package CYANA 3.97 (311). During the process of automatic NOESY assignment, CYANA assigned 96.3% of all cross peaks. The final structure was calculated using 66 interproton distance restraints, 6 disulfide-bond restraints and 23 dihedral-angle restraints. 200 structures were calculated, then the 20 best structures (chosen in the basis of the final CYANA target function values and stereochemical quality as judged by Molprobity) were used to represent the solution structure of Ate1a. The structural ensemble is available from the Protein Data Bank (accession code 6AZA).

4.4.7 Activity of Ate1a

4.4.7.1 Electrophysiological characterisation of Ate1a

Two-electrode voltage-clamp oocyte electrophysiology

The pharmacological effect of Ate1a was analysed by heterologous expression of $rK_V1.1$,

rK$_V$1.2, hK$_V$1.3, rK$_V$1.4, rK$_V$1.5, rK$_V$1.6, Shaker IR, rK$_V$2.1, hK$_V$3.1, rK$_V$4.2, K$_V$7.2, K$_V$11.1, rNa$_V$1.2, rNa$_V$1.3, rNa$_V$1.4, hNa$_V$1.5, mNa$_V$1.6, hNa$_V$1.7, rNa$_V$1.8, rASIC1a, rASIC1b, rASIC2a, and rASIC3 in *Xenopus laevis* oocytes. The linearized plasmids were transcribed using the T7 or SP6 mMESSAGE-mMACHINE transcription kit (Ambion, Waltham, MA, USA). The K$_V$1.1 triple mutant channel was constructed as previously describes (312). Oocytes were injected with 50 nL of cRNA at a concentration of 0.05–1 ng/nL using a micro-injector (Drummond Scientific, Broomall, PA, USA). The oocytes were maintained in an ND96 solution (in mM: 96 NaCl, 2 KCl, 1.8 $CaCl_2$, 2 $MgCl_2$ and 5 HEPES; pH 7.4), supplemented with 50 μg/mL gentamicin sulfate.

Two-electrode voltage-clamp recordings were performed at room temperature (18–22 °C) using a Geneclamp 500 or Axoclamp 900A amplifier (Molecular Devices, Sunnyvale, CA, USA) controlled by a pClamp data acquisition system (Axon Instruments, Union City, CA, USA). Whole cell currents from oocytes were recorded 1–5 days after injection. The bath solution composition was ND96 (in mM: 2 NaCl, 96 KCl, 1.8 $CaCl_2$, 2 $MgCl_2$ and 5 HEPES; pH 7.4). Voltage and current electrodes were filled with KCl (3 M). The resistance of both electrodes was kept between 0.8 and 1.0 MΩ. K$_V$ currents were filtered at 0.5 or 2 kHz using a four-pole low-pass Bessel filter, and leak subtraction was performed using a −P/4 protocol. K$_V$1.1−K$_V$1.6 and Shaker IR currents were evoked by 500 ms depolarization to 0 mV followed by a 500 ms pulse to −50 mV, from a holding potential of −90 mV. K$_V$2.1, K$_V$3.1, K$_V$4.2 and K$_V$4.3 currents were elicited by 500 ms pulses to +20 mV from a holding potential of −90 mV. Current traces of *h*ERG channels were elicited by applying a +40 mV pulse for 2.5 s followed by a step to −120 mV for 2.5 s. Na$_V$ current were evoked by 100 ms depolarization pulse from a holding potential of −90 mV to −20 mV, with the exception of Na$_V$1.8 which were pulsed to 0 mV. Na$_V$ data were digitized at 20 kHz; leak and background conductance were identified by blocking channels with tetrodotoxin and subtracted from currents. ASIC currents were acquired (digitized 2 kHz and filtered at 0.01 Hz) and elicited by a drop in pH from 7.45 to 6.5, 5.5, 4.5, and 6.3 (for ASIC1a, ASIC1b, ASIC2a, and ASIC3, respectively).

Manual patch clamp electrophysiology

Due to the lack of a functional Na$_V$1.1 clone for oocyte screening, activity was assessed by manual patch clamp electrophysiology. HEK293 cells heterologously expressing hNa$_V$1.1 (SB Drug Discovery, Glasgow, UK) were maintained at 37 °C in a humidified 5% CO_2 incubator in minimal essential medium supplemented with 10% FBS v/v, 2 mM L-glutamine and

selection antibiotics as recommended by the manufacturer. Cells were grown to 70–80% confluence and passaged every 2–4 days using Detachin (Genlantis, San Diego, CA, USA).

Whole cell patch-clamp recordings were obtained using a MultiClamp 700B amplifier (Molecular Devices, Sunnyvale, CA, USA). Patch pipettes were pulled from standard wall borosilicate glass capillaries (1.5 mm x 0.86 mm, OD/ID; SDR Scientific, Sydney, AUS) using a microelectrode puller (P-97; Sutter Instrument Co., Novato, CA, USA) and had a resistance of 1.2–1.5 MW when filled with pipette solution. The pipette solution was composed of (in mM) 150 CsCl, 1 EGTA, 10 HEPES and was adjusted to pH 7.2 with CsOH. The external bath solution consisted of (in mM) 140 NaCl, 4 KCl, 1 $MgCl_2$, 2 $CaCl_2$, 10 HEPES and adjusted to pH 7.4 with NaOH. Currents were monitored for at least 5 min after establishing whole-cell configuration to allow currents to stabilize. The pulse protocol consisted of cells being held at -90 mV for 10 s, followed by a hyperpolarizing step to -120 mV for 200 ms, then a depolarizing step to -15 mV for 50 ms. Series resistance and prediction compensation between 50–75% was applied to reduce voltage errors. Recorded currents were acquired with a Digidata 1550B (Molecular Devices) converter at 50 kHz after passing through a low-pass Bessel filter of 10 kHz. A P/6 subtraction protocol provided by the Clampex (Molecular Devices) acquisition software was used to remove linear leak and residual capacitance artifacts.

Data analysis

To assess the Ate1a concentration–response relationships, data were fitted with the Hill equation $y = 100/[1 + (IC_{50}/[\text{toxin}])h]$, where y is the amplitude of the toxin-induced effect, IC_{50} is the toxin concentration at half-maximal efficacy, [toxin] is the toxin concentration and h is the Hill coefficient. In order to investigate the current–voltage (I–V) relationship, current traces were evoked by 10 mV depolarization steps from a holding potential of −90 mV. The values of I_K were plotted as function of voltage and fitted using the Boltzmann equation $I_K/I_{max} = [1 + \exp(V_g - V)/k]^{-1}$, where I_{max} represents maximal I_K, V_g is the voltage corresponding to half-maximal current and k is the slope factor. To assess the concentration dependence of the Ate1a induced inhibitory effects, a concentration–response curve was constructed in which the percentage of current inhibition was plotted as a function of toxin concentration. Data were fitted with the Hill equation. All data represent at least three independent experiments ($n \geq 3$) and are presented as mean ± standard error. Comparison of two sample means was made using a paired Student's t test ($P < 0.05$). All data were analyzed using clampfit 10.3 (Molecular Devices) and origin 7.5 software (Origin Lab., Northampton,

MA, USA).

4.4.7.2 Antimicrobial Assays

Antimicrobial screening was performed by the Community for Antimicrobial Drug Discovery (CO-ADD; www.co-add.org). Ate1a was prepared at 10 mg/mL in DMSO and diluted in Mueller Hinton Broth (MHB; BD, Cat. No.211443). Initial antimicrobial screening was conducted by whole-cell growth inhibition assays using a single Ate1a concentration of 256 μg/mL, in duplicate. Inhibition of growth was measured against five bacteria: *Escherichia coli* (ATCC 25922), *Klebsiella pneumoniae* (ATCC 700603), *Acinetobacter baumannii* (ATCC 19606), *Pseudomonas aeruginosa* (ATCC 27853) and *Staphylococcus aureus* (MRSA; ATCC 43300), and two fungi: *Candida albicans* (ATCC 90028) and *Cryptococcus neoformans* (ATCC 208821). Dose-response was also examined with two-fold across the wells of 96-well non-binding surface (NBS) plates (Corning; Cat. No. 3641) in duplicate.

Bacterial strains were cultured in MHB at 37°C overnight, then diluted 40-fold and incubated at 37°C for a further 2–3 h. The resultant mid-log phase cultures were diluted in MHB and added to each well of the compound-containing 96-well plates to give a final cell density of $5x10^5$ CFU/mL. The final Ate1a concentration range was 256 – 0.125 μg/mL. The plates were covered and incubated at 37°C for 20 h. Optical density was read at 600 nm (OD_{600}) using a Tecan M1000 Pro Spectrophotometer. Minimal inhibitory concentration (MIC) was determined as the lowest concentration corresponding to ≥95% growth inhibition, relative to a growth control, as measured from the OD_{600}.

For antifungal assays Ate1a was serially diluted in Yeast Nitrogen Broth (YNB) two-fold across the wells of 96-well non-binding surface (NBS) plates (Corning; Cat. No. 3641) plated in duplicate. *Candida albicans* and *Cryptococcus neoformans* were cultured on YPD agar at 35°C for 24 and 48 h respectively. A minimum of five single colonies were taken from each agar plate and dissolved in sterile water; the solution was then adjusted to OD_{530} = 0.3. The solution was diluted in YNB and added to each well of the Ate1a-containing 96-well plates giving a final cell density of $2.5x10^3$ CFU/mL and a final Ate1a concentration of 256–0.125 μg/mL. The plates were covered and incubated at 35°C for 36 h. MIC was the lowest concentration with ≥85% growth inhibition and was determined by optical density after 36 h incubation using a Biotek Synergy HTx Plate reader following at OD_{630} for *Candida albicans* and at $OD_{570-600}$ for *Cryptococcus neoformans*.

4.4.7.3 Interactions of Ate1a with lipid bilayers

Surface plasmon resonance (SPR) was used to monitor the affinity of Ate1a for lipid membranes using a Biacore 3000 instrument (GE healthcare) at 25ºC. Synthetic POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) and POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine) (Avanti polar lipids) were used to prepare small unilamellar vesicles (SUVs, 50 nm diameter) composed of POPC or POPC/POPS (4:1 molar ratio) dispersed in HEPES buffer (10 mM HEPES containing 150 mM NaCl, pH 7.4) and homogenized by extrusion. SUVs were deposited onto an L1 chip for 40 min at a flow rate of 2 μL/min. Serial two-fold dilutions of Ate1a, starting from 64 μM, were injected over deposited lipid bilayers for 180 s at a flow rate of 5 μL/min (association phase); dissociation was followed for 600 s (313, 314). An N-to-C cyclized version of gomesin was included for comparison. The chip was regenerated as before (315). All solutions were freshly prepared and filtered using a 0.22 μm filter; HEPES buffer was used as running buffer. Response units were normalized to peptide-to-lipid ratio (P/L) as previously described (313).

4.4.7.4 Hemolysis studies

Erythrocytes were isolated from fresh human blood collected from three healthy donors using protocols approved by the Human Research Ethics Committee at The University of Queensland. Erythrocytes were resuspended at 0.25% (v/v) in phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) and incubated with two-fold dilutions of Ate1a for 1 h at 37ºC. Hemolysis was quantified as described (316) by measuring hemoglobin release at 405 nm. Melittin and cyclic gomesin were used as controls.

4.4.7.5 Cell culture and cell viability assays

HeLa, MCF-7, HFF-1 were grown in DMEM supplemented with 1% penicillin/streptomycin and 10% v/v (HeLa, MCF-7) or 15% v/v (HFF-1) FBS. Cells were incubated in a humidified atmosphere (5% CO_2, 37ºC) and split by dilution after reaching confluency. Toxicity of Ate1a was determined using a resazurin colorimetric assay, as described (317). Briefly, cells were seeded in a 96-well plate ($5x10^3$ cells/well) the day before the assay. Ate1a was diluted in

medium without serum (two-fold dilutions starting at 64 μM) and added to cells; resazurin (0.05% w/v) was added 2 h after the peptide and incubated for another 24 h. Fluorescence emission intensity (excitation 565 nm, emission 584 nm) was measured using a plate reader (Tecan M1000 Pro Spectrophotometer). Melittin and cyclic gomesin were included as controls, and samples with buffer or Triton X-100 (0.1% v/v) were included to estimate 100% and 0% viability, respectively. Assays were repeated at least three times.

4.4.7.6 Toxicity bioassays

Toxicity of Ate1a to brine shrimps (*A. salina*) and amphipods (family Talitridae) was examined as previously described (18, 318). For assays where the toxin was dissolved in medium, synthetic Ate1a was dissolved to a concentration of 0.5 mg/mL in filtered artificial seawater. Assays were performed in 24-well plates for shrimps and 6-well plates for amphipods (Thermo-Fisher Scientific). Paralysis and lethality were assessed by microscopic observation and responsiveness to contact with a plastic tip. Bovine serum albumin (5 mg/mL) was used as a control (no toxicity). For injection assays, Ate1a was diluted with a physiological solution for crustaceans (in mM: NaCl 470.4, KCl 8.0, $CaCl_2$ 18.0, $MgCl_2$1.5, $NaHCO_3$ 6.0 and glucose 5.6). The injection volume was 9.4 nL. Groups of ten amphipods (4.35–11.53 mg) were challenged with 5.3 mM of toxin and observed for mortality or paralysis up to 4 h.

CHAPTER 5
Conclusions and future directions

The ultimate aim of this thesis was to provide the first comprehensive insight into the composition and biodiscovery potential of sea anemone venoms. Through a uniquely multifaceted approach we studied multiple aspects of these venoms. As discussed below, together this thesis gives the to date most comprehensive insight into one of the oldest extant venom systems known, and is thereby a substantial contribution to the field of venomics.

5.1 Sea anemone venom composition

Previously published works were using mainly transcriptomic and genomic approaches to access sea anemone venoms composition. However, homology-based annotation of toxins from transcriptomic and genomic data is insufficient for providing a full picture of the venom arsenal as it is inherently limited to finding candidates with significant sequence homology to known toxins, and it is prone to false positives by the fact that toxins typically evolve from proteins with functions unrelated to envenomation. In addition, this is particularly problematic for cnidarians because transcriptome data obtained from tentacles or whole body contains transcripts from tissues not specifically involved in venom production. In this thesis, we highlighted the importance of using a combined transcriptomics/proteomics approach to provide a holistic overview of the complexity of the venom arsenal of sea anemones. Even though this project was likely to discover novel venom components, it was surprising that the application of this approach to a representative of the well-studied genus *Stichodactyla* uncovered twelve new families of putative toxins. These new families contained no known structural or functional domains nor showed any sequence homology to any characterized protein, highlighting how much there is still to be learnt about sea anemone venoms.

Another surprise was the differences between purely transcriptome and proteomics analyses. Previously published works using only transcriptomic, usually use transcript expression levels to identify putative toxins based on the observation that toxin genes are more highly expressed in venom glands that are actively engaged in the process of venom regeneration compared to those that are replete. However, cnidarians represent the only lineage of venomous animals that lack a centralized venom system and the process of venom regeneration remains largely uncharacterized. Our results showed poor correlation between toxin transcript levels and abundance in venom. Altogether, this thesis highlighted the importance of using proteomics of milked venom to correctly identify venom proteins/peptides, both known and novel, while

minimizing false positive identifications from non-toxin homologues identified in transcriptomes of venom-producing tissues.

Another benefit from that this thesis brought, is the improvement in the sea anemone toxins database in the protein level. Transcriptome annotation is limited by the lack of a comprehensive cnidarian proteomics database to match against the transcripts. In addition, final matching of proteome and transcriptome data requires multiple bioinformatics tools. We developed a pipeline using Galaxy that contains all tools required for these analyse.

5.2 Biodiscovery potential of sea anemone venoms

Sea anemone venoms have long been recognised as a rich source of peptides with diverse structure and pharmacological properties, but they still contain many uncharacterised bioactive compounds with novel scaffolds. While the majority of studies target medically-important or abundant organisms for ethical or practical reasons, research on rare organisms or less-studied species often provides a promising source of novelty. As described in chapter 3, here we reported the structure, activity, function and evolution of a new, sixth type of sea anemone K_V toxin. κ-Actitoxin-Ate1a (henceforth Ate1a), from venom of the Waratah sea anemone *Actinia tenebrosa*. Comprised of just 17 residues, Ate1a is the shortest sea anemone toxin reported to date, and it adopts a novel three-dimensional structure that we named the Proline-Hinged Asymmetric β-hairpin (PHAB) fold.

Aiming to understand the function of Ate1a, we used a combination of toxicity assays and a wide molecular target screening. Ate1a lacks antimicrobial activity and instead serves a predatory function via potent inhibition of prey K_V channels. Although the pharmacological activity of a toxin can provide clues to its ecological function, it is not by itself definitive. However, the near-universal distribution of nematocytes in the epithelium of sea anemones means that toxin function can be inferred from tissue distribution. We therefore investigated the tissue distribution of Ate1a using MALDI-MS imaging (MSI). MSI revealed that Ate1a is non-uniformly distributed within the body of *A. tenebrosa*, with almost exclusive localization in tentacles, suggesting that it is involved in prey capture. IMS was a powerful technique that, in combination with the toxicity assays and molecular target, gave us insights about the ecological role of Ate1a.

We also searched for the presence of analogues of Ate1a in other sea anemones. This family of toxin has evolved in a particular way. While most families of predatory toxins evolve via

bursts of extensive duplication and diversification, this is not the case for the Ate1a toxin family, whose members remain remarkably well conserved despite their ancient evolution. Our data suggest that this extreme conservation is due to intra-gene concerted evolution, a process that has not been previously reported for any toxin family and which we propose is a hitherto unrecognised mechanism of maintaining efficient secretion of ecologically important toxins.

Ate1a is only one example of how the combination of multiple techniques is essential for fully characterization of a toxin. Similar approach should be applied for other toxins to gather enough data to give us insights of how the combination of these toxins work together in the process of envenomation. This is essential for understanding of the ecological role of sea anemone venoms.

5.3 Future directions

While the works presented herein provides a first holistic insight into a sea anemone venom and a full characterization of a new scaffold (Ate1a), it also raises a large number of questions and prospects for future studies. This thesis has uncovered a vast number of proteins and peptides that still need structure and functional characterization to provide insight into their roles in sea anemone venoms. Moreover, the venom of other species remained unexplored and it will be interesting to see the outcomes of future studies addressing this shortcoming of our current understanding of sea anemone venom composition, ecological role and the evolutionary forces that shape it.

5.4 Conclusions

Although there is still much to learn about the composition of sea anemone venoms and the role of individual venom components in prey capture, defence and intraspecific competition, this thesis lays the foundation for uncovering the role of individual toxins in sea anemone venom and how they contribute to the envenomation of prey, predators, and competitors.

References

1. Technau U & Schwaiger M (2015) Recent advances in genomics and transcriptomics of cnidarians. *Mar Genomics* 24 Pt 2:131-138.
2. Rodriguez E*, et al.* (2014) Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of hexacorals. *PLoS One* 9(5):e96998.
3. Brusca RC, Brusca GJ, & Haver N (2003) *Invertebrates* (Sinauer Associates, Sunderland, Mass.) 2nd Ed pp xix, 936 p.
4. Stewart ZK, Pavasovic A, Hock DH, & Prentis PJ (2017) Transcriptomic investigation of wound healing and regeneration in the cnidarian Calliactis polypus. *Sci Rep-Uk* 7.
5. Galliot B & Schmid V (2002) Cnidarians as a model system for understanding evolution and regeneration. *Int J Dev Biol* 46(1):39-48.
6. Rodriguez E*, et al.* (2014) Hidden among Sea Anemones: The First Comprehensive Phylogenetic Reconstruction of the Order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) Reveals a Novel Group of Hexacorals. *Plos One* 9(5).
7. Chintiroglou C & Koukouras A (1992) The Feeding-Habits of 3 Mediterranean-Sea Anemone Species, Anemonia-Viridis (Forskal), Actinia-Equina (Linnaeus) and Cereus-Pedunculatus (Pennant). *Helgolander Meeresun* 46(1):53-68.
8. Ruppert EE, Fox RS, & Barnes RD (2004) *Invertebrate zoology : a functional evolutionary approach* (Thomson-Brooks/Cole, Belmont, CA) 7th Ed pp xvii, 963, 926 p.
9. Dubinsky Z & Stambler N (2011) *Coral reefs : an ecosystem in transition* (Springer, Dordrecht) pp ix, 552 pages.
10. Moran Y*, et al.* (2012) Neurotoxin localization to ectodermal gland cells uncovers an alternative mechanism of venom delivery in sea anemones. *Proc. Biol. Sci.* 279(1732):1351–1358.
11. Reft AJ & Daly M (2012) Morphology, distribution, and evolution of apical structure of nematocysts in hexacorallia. *J Morphol* 273(2):121-136.
12. David CN*, et al.* (2008) Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trends in genetics : TIG* 24(9):431-438.
13. Nuchter T, Benoit M, Engel U, Ozbek S, & Holstein TW (2006) Nanosecond-scale kinetics of nematocyst discharge. *Curr Biol* 16(9):R316-R318.
14. Kass-Simon G & Scappaticci AA (2002) The behavioral and developmental physiology of nematocysts. *Can J Zool* 80(10):1772-1794.
15. Beckmann A & Ozbek S (2012) The nematocyst: a molecular map of the cnidarian stinging organelle. *Int J Dev Biol* 56(6-8):577-582.
16. Daly M (2017) *Functional and Genetic Diversity of Toxins in Sea Anemones* pp 87-104.
17. Basulto A*, et al.* (2006) Immunohistochemical targeting of sea anemone cytolysins on tentacles, mesenteric filaments and isolated nematocysts of Stichodactyla helianthus. *J Exp Zool A Comp Exp Biol* 305(3):253-258.

18. Moran Y*, et al.* (2012) Neurotoxin localization to ectodermal gland cells uncovers an alternative mechanism of venom delivery in sea anemones. *Proceedings. Biological Sciences* 279(1732):1351–1358.

19. Ardelean A & Fautin DG (2004) A new species of the sea anemone Megalactis (Cnidaria : Anthozoa : Actiniaria : Actinodendridae) from Taiwan and designation of a neotype for the type species of the genus. *P Biol Soc Wash* 117(4):488-504.

20. Macrander J, Broe M, & Daly M (2016) Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemones. *Genome Biol Evol* 8(8):2358-2375.

21. Kitatani R, Yamada M, Kamio M, & Nagai H (2015) Length Is Associated with Pain: Jellyfish with Painful Sting Have Longer Nematocyst Tubules than Harmless Jellyfish. *Plos One* 10(8).

22. Shick JM (1991) *A functional biology of sea anemones* (Chapman & Hall, London) pp xxi, 395p.

23. Logashina YA*, et al.* (2017) New disulfide-stabilized fold provides sea anemone peptide to exhibit both antimicrobial and TRPA1 potentiating properties. *Toxins (Basel)* 9(5).

24. Osmakov DI*, et al.* (2013) Sea anemone peptide with uncommon β-hairpin structure inhibits acid-sensing ion channel 3 (ASIC3) and reveals analgesic activity. *J. Biol. Chem.* 288(32):23116–23127.

25. Nevalainen TJ*, et al.* (2004) Phospholipase A_2 in cnidaria. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 139(4):731-735.

26. Anderluh G & Macek P (2002) Cytolytic peptide and protein toxins from sea anemones (Anthozoa: Actiniaria). *Toxicon* 40(2):111-124.

27. Monastyrnaya M*, et al.* (2016) Kunitz-Type Peptide HCRG21 from the Sea Anemone Heteractis crispa Is a Full Antagonist of the TRPV1 Receptor. *Mar Drugs* 14(12).

28. Andreev YA*, et al.* (2008) Analgesic compound from sea anemone Heteractis crispa is the first polypeptide inhibitor of vanilloid receptor 1 (TRPV1). *J Biol Chem* 283(35):23914-23921.

29. Cuypers E, Peigneur S, Debaveye S, Shiomi K, & Tytgat J (2011) TRPV1 Channel as New Target for Marine Toxins: Example of Gigantoxin I, a Sea Anemone Toxin Acting Via Modulation of the PLA_2 Pathway. *Acta Chim Slov* 58(4):735-741.

30. Norton RS (1991) Structure and Structure-Function-Relationships of Sea-Anemone Proteins That Interact with the Sodium-Channel. *Toxicon* 29(9):1051-1084.

31. Honma T & Shiomi K (2006) Peptide toxins in sea anemones: structural and functional aspects. *Marine biotechnology* 8(1):1-10.

32. Kozlov S & Grishin E (2012) Convenient nomenclature of cysteine-rich polypeptide toxins from sea anemones. *Peptides* 33(2):240-244.

33. Oliveira JS, Fuentes-Silva D, & King GF (2012) Development of a rational nomenclature for naming peptide and protein toxins from sea anemones. *Toxicon* 60(4):539-550.

34. Norton RS (2013) Sea Anemone Peptides. *Handbook of Biologically Active Peptides*, eds Ribeiro SM, Porto WF, Silva ON, Santos MdO, Dias SC, & Franco OL (Elsevier), Second Ed, pp 430–436.

35. King GF, Gentz MC, Escoubas P, & Nicholson GM (2008) A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* 52(2):264-276.

36. Undheim EAB*, et al.* (2014) Clawing through Evolution: Toxin Diversification and Convergence in the Ancient Lineage Chilopoda (Centipedes). *Mol Biol Evol* 31(8):2124-2148.

37. Fry BG*, et al.* (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annual Review of Genomics and Human Genetics* 10:483-511.

38. Sunagar K*, et al.* (2013) Three-fingered RAVERs: Rapid Accumulation of Variations in Exposed Residues of snake venom toxins. *Toxins (Basel)* 5(11):2172-2208.

39. de Freitas JC & Sawaya MI (1986) Anomalies in sea-urchin egg development induced by a novel purine isolated from the sea-anemone *Bunodosoma caissarum*. *Toxicon* 24(8):751-755.

40. de Freitas JC & Sawaya MI (1990) Increase of mammalian intestinal motility by the iminopurine caissarone isolated from the sea anemone *Bunodosoma caissarum*. *Toxicon* 28(9):1029-1037.

41. Garateix A*, et al.* (1996) Antagonism of glutamate receptors by a chromatographic fraction from the exudate of the sea anemone Phyllactis flosculifera. *Toxicon* 34(4):443-450.

42. Murakami M*, et al.* (2014) Emerging roles of secreted phospholipase A enzymes: The 3rd edition. *Biochimie*.

43. Kordis D (2011) Evolution of phospholipase A_2 toxins in venomous animals. *Acta Chim Slov* 58(4):638-646.

44. Parker MW & Feil SC (2005) Pore-forming protein toxins: from structure to function. *Progress in biophysics and molecular biology* 88(1):91-142.

45. Elliott RC, Konya RS, & Vickneshwara K (1986) The isolation of a toxin from the dahlia sea anemone, Tealia felina L. *Toxicon* 24(2):117-122.

46. Zykova TA, Monastyrnaia MM, Apalikova OV, Shvets TV, & Kozlovskaia EP (1998). Low-molecular cytolysins and trypsin inhibitors from sea anemone *Radianthus macrodactylus*. Isolation and partial characterization. *Bioorg Khim* 24(7):509-516.

47. Monastyrnaya M*, et al.* (2010) Actinoporins from the sea anemones, tropical *Radianthus macrodactylus* and northern *Oulactis orientalis*: Comparative analysis of structure-function relationships. *Toxicon* 56(8):1299-1314.

48. Suput D (2009) *In vivo* effects of cnidarian toxins and venoms. *Toxicon* 54(8):1190-1200.

49. Alvarez C*, et al.* (2009) Sticholysins, two pore-forming toxins produced by the Caribbean Sea anemone *Stichodactyla helianthus*: their interaction with membranes. *Toxicon* 54(8):1135-1147.

50. Anderluh G, Sepcic K, Turk T, & Macek P (2011) Cytolytic proteins from cnidarians - an overview. *Acta Chim Slov* 58(4):724-729.

51. Wang Y, Yap LL, Chua KL, & Khoo HE (2008) A multigene family of Heteractis magnificalysins (HMgs). *Toxicon* 51(8):1374-1382.

52. Garcia-Ortega L*, et al.* (2011) The behavior of sea anemone actinoporins at the water-membrane interface. *Biochimica et biophysica acta* 1808(9):2275-2288.

53. Moran Y*, et al.* (2013) Analysis of soluble protein contents from the nematocysts of a model sea anemone sheds light on venom evolution. *Mar Biotechnol (NY)* 15(3):329-339.

54. Rachamim T*, et al.* (2015) The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan. *Mol Biol Evol* 32(3):740-753.

55. Valle A*, et al.* (2015) The multigene families of actinoporins (part I): Isoforms and genetic structure. *Toxicon* 103:176-187.

56. Basulto A*, et al.* (2006) Immunohistochemical targeting of sea anemone cytolysins on tentacles, mesenteric filaments and isolated nematocysts of *Stichodactyla helianthus*. *Journal of Experimental Zoology Part A, Comparative Experimental Biology* 305(3):253-258.

57. Garcia-Linares S*, et al.* (2013) Three-dimensional structure of the actinoporin sticholysin I. Influence of long-distance effects on protein function. *Arch Biochem Biophys* 532(1):39-45.

58. Morante K, Caaveiro JM, Viguera AR, Tsumoto K, & Gonzalez-Manas JM (2015) Functional characterization of Val60, a key residue involved in the membrane-oligomerization of fragaceatoxin C, an actinoporin from *Actinia fragacea*. *FEBS Lett* 589(15):1840-1846.

59. Bakrac B*, et al.* (2008) Molecular determinants of sphingomyelin specificity of a eukaryotic pore-forming toxin. *J Biol Chem* 283(27):18665-18677.

60. Garcia-Linares S*, et al.* (2014) The sea anemone actinoporin (Arg-Gly-Asp) conserved motif is involved in maintaining the competent oligomerization state of these pore-forming toxins. *FEBS J* 281(5):1465-1478.

61. Malovrh P*, et al.* (2000) Structure-function studies of tryptophan mutants of equinatoxin II, a sea anemone pore-forming protein. *The Biochemical journal* 346 Pt 1:223-232.

62. Hong Q*, et al.* (2002) Two-step membrane binding by Equinatoxin II, a pore-forming toxin from the sea anemone, involves an exposed aromatic cluster and a flexible helix. *J Biol Chem* 277(44):41916-41924.

63. Mancheno JM, Martin-Benito J, Martinez-Ripoll M, Gavilanes JG, & Hermoso JA (2003) Crystal and electron microscopy structures of sticholysin II actinoporin reveal insights into the mechanism of membrane pore formation. *Structure* 11(11):1319-1328.

64. Malovrh P*, et al.* (2003) A novel mechanism of pore formation: membrane penetration by the N-terminal amphipathic region of equinatoxin. *J Biol Chem* 278(25):22678-22685.

65. Gutierrez-Aguirre I*, et al.* (2004) Membrane insertion of the N-terminal alpha-helix of equinatoxin II, a sea anemone cytolytic toxin. *The Biochemical journal* 384(Pt 2):421-

428.

66. Belmonte G, Pederzolli C, Macek P, & Menestrina G (1993) Pore Formation by the Sea-Anemone Cytolysin Equinatoxin-Ii in Red-Blood-Cells and Model Lipid-Membranes. *J Membrane Biol* 131(1):11-22.

67. Tejuca M, Serra MD, Ferreras M, Lanio ME, & Menestrina G (1996) Mechanism of membrane permeabilization by sticholysin I, a cytolysin isolated from the venom of the sea anemone *Stichodactyla helianthus*. *Biochemistry* 35(47):14947-14957.

68. Alegre-Cebollada J, Martinez del Pozo A, Gavilanes JG, & Goormaghtigh E (2007) Infrared spectroscopy study on the conformational changes leading to pore formation of the toxin sticholysin II. *Biophysical journal* 93(9):3191-3201.

69. Anderluh G & Macek P (2003) Dissecting the actinoporin pore-forming mechanism. *Structure* 11(11):1312-1313.

70. Valcarcel CA*, et al.* (2001) Effects of lipid composition on membrane permeabilization by sticholysin I and II, two cytolysins of the sea anemone *Stichodactyla helianthus*. *Biophysical journal* 80(6):2761-2774.

71. Tejuca M, Anderluh G, & Dalla Serra M (2009) Sea anemone cytolysins as toxic components of immunotoxins. *Toxicon* 54(8):1206-1214.

72. Avila AD, Mateo de Acosta C, & Lage A (1989) A carcinoembryonic antigen-directed immunotoxin built by linking a monoclonal antibody to a hemolytic toxin. *Int J Cancer* 43(5):926-929.

73. Potrich C*, et al.* (2005) Cytotoxic activity of a tumor protease-activated pore-forming toxin. *Bioconjugate chemistry* 16(2):369-376.

74. Tejuca M*, et al.* (2004) Construction of an immunotoxin with the pore forming protein StI and ior C5, a monoclonal antibody against a colon cancer cell line. *International immunopharmacology* 4(6):731-744.

75. Thrush GR, Lark LR, Clinchy BC, & Vitetta ES (1996) Immunotoxins: an update. *Annual review of immunology* 14:49-71.

76. Ghetie V & Vitetta ES (2001) Chemical construction of immunotoxins. *Molecular biotechnology* 18(3):251-268.

77. Lenhoff DAHHM (1973) Assay and properties of the hemolysis activity of pure venom from the nematocysts of the acontia of the sea anemone *Aiptasia pallida*. *Archives of Biochemistry and Biophysics* 159:629-638.

78. Razpotnik A, Krizaj I, Kem WR, Macek P, & Turk T (2009) A new cytolytic protein from the sea anemone *Urticina crassicornis* that binds to cholesterol- and sphingomyelin-rich membranes. *Toxicon* 53(7-8):762-769.

79. Cline EI, Wiebe LI, Young JD, & Samuel J (1995) Toxic effects of the novel protein UpI from the sea anemone *Urticina piscivora*. *Pharmacol Res* 32(5):309-314.

80. Hessinger DA & Lenhoff HM (1973) Binding of active and inactive hemolytic factor of sea anemone nematocyst venom to red blood cells. *Biochem Biophys Res Commun* 53(2):475-481.

81. Avigad AWBLS (1978) A cholesterol-inhibitable cytolytic from the sea anemone

Metridium senile. *Biochimica et Biophysica Acta (BBA) - General Subjects* 541(1):96-106.

82. Bernheimer AW, Avigad LS, & Kim K (1979) Comparison fo metridiolysin from the sea anemone with thiol-activated cytolysins from bacteria. *Toxicon* 17(1):69-75.

83. Muller-Eberhard HJ (1986) The membrane attack complex of complement. *Annual review of immunology* 4:503-528.

84. Liu CC, Walsh CM, & Young JD (1995) Perforin: structure and function. *Immunology today* 16(4):194-201.

85. Putnam NH*, et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317(5834):86-94.

86. Rees B & Bilwes A (1993) Three-dimensional structures of neurotoxins and cardiotoxins. *Chemical research in toxicology* 6(4):385-406.

87. Cristofori-Armstrong B & Rash LD (2017) Acid-sensing ion channel (ASIC) structure and function: Insights from spider, snake and sea anemone venoms. *Neuropharmacology*.

88. Rodriguez AA*, et al.* (2014) A novel sea anemone peptide that inhibits acid-sensing ion channels. *Peptides* 53:3–12.

89. Moran Y, Gordon D, & Gurevitz M (2009) Sea anemone toxins affecting voltage-gated sodium channels — molecular and evolutionary features. *Toxicon* 54(8):1089-1101.

90. Jouiaei M*, et al.* (2015) Evolution of an ancient venom: recognition of a novel family of cnidarian toxins and the common evolutionary origin of sodium and potassium neurotoxins in sea anemone. *Mol Biol Evol* 32(6):1598-1610.

91. Bosmans F & Tytgat J (2007) Sea anemone venom as a source of insecticidal peptides acting on voltage-gated Na^+ channels. *Toxicon* 49(4):550-560.

92. Deuis JR, Mueller A, Israel MR, & Vetter I (2017) The pharmacology of voltage-gated sodium channel activators. *Neuropharmacology*.

93. Diochot S & Lazdunski M (2009) Sea anemone toxins affecting potassium channels. *Prog Mol Subcell Biol* 46:99-122.

94. Orts DJB*, et al.* (2013) BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker. *FEBS J.* 280(19):4839–4852.

95. Castaneda O & Harvey AL (2009) Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels. *Toxicon* 54(8):1119-1124.

96. The UniProt C (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45(D1):D158-D169.

97. Manoleras N & Norton RS (1994) Three-dimensional structure in solution of neurotoxin III from the sea anemone *Anemonia sulcata*. *Biochemistry* 33(37):11051-11061.

98. Beress L, Beress R, & Wunderer G (1975) Isolation and characterisation of three polypeptides with neurotoxic activity from *Anemonia sulcata*. *FEBS Lett* 50(3):311-314.

99. Schweitz H*, et al.* (1981) Purification and pharmacological properties of eight sea

anemone toxins from *Anemonia sulcata*, *Anthopleura xanthogrammica*, *Stoichactis giganteus*, and *Actinodendron plumosum*. *Biochemistry* 20(18):5245-5252.

100. Moran Y*, et al.* (2007) Molecular analysis of the sea anemone toxin Av3 reveals selectivity to insects and demonstrates the heterogeneity of receptor site-3 on voltage-gated Na^+ channels. *The Biochemical journal* 406(1):41-48.

101. Suarez-Carmona M, Hubert P, Delvenne P, & Herfs M (2015) Defensins: "Simple" antimicrobial peptides or broad-spectrum molecules? *Cytokine Growth Factor Rev* 26(3):361-370.

102. Shafee TM, Lay FT, Hulett MD, & Anderson MA (2016) The Defensins Consist of Two Independent, Convergent Protein Superfamilies. *Mol Biol Evol* 33(9):2345-2356.

103. Chagot B, Diochot S, Pimentel C, Lazdunski M, & Darbon H (2005) Solution structure of APETx1 from the sea anemone *Anthopleura elegantissima*: a new fold for an HERG toxin. *Proteins* 59(2):380-386.

104. Chagot B*, et al.* (2005) Solution structure of APETx2, a specific peptide inhibitor of ASIC3 proton-gated channels. *Protein Sci* 14(8):2003-2010.

105. Smith JJ & Blumenthal KM (2007) Site-3 sea anemone toxins: molecular probes of gating mechanisms in voltage-dependent sodium channels. *Toxicon* 49(2):159-170.

106. Diochot S*, et al.* (2004) A new sea anemone peptide, APETx2, inhibits ASIC3, a major acid-sensitive channel in sensory neurons. *The EMBO journal* 23(7):1516-1525.

107. Diochot S, Loret E, Bruhn T, Beress L, & Lazdunski M (2003) APETx1, a new toxin from the sea anemone *Anthopleura elegantissima*, blocks voltage-gated human ether-a-go-go-related gene potassium channels. *Molecular pharmacology* 64(1):59-69.

108. Diochot S, Schweitz H, Beress L, & Lazdunski M (1998) Sea anemone peptides with a specific blocking activity against the fast inactivating potassium channel $K_V3.4$. *J Biol Chem* 273(12):6744-6749.

109. Liu P, Jo S, & Bean BP (2012) Modulation of neuronal sodium channels by the sea anemone peptide BDS-I. *Journal of neurophysiology* 107(11):3155-3167.

110. Peigneur S*, et al.* (2012) A natural point mutation changes both target selectivity and mechanism of action of sea anemone toxins. *Faseb J* 26(12):5141-5151.

111. Beress L*, et al.* (1985) DE 3324689 A1.

112. Yeung SY, Thompson D, Wang Z, Fedida D, & Robertson B (2005) Modulation of K_V3 subfamily potassium currents by the sea anemone toxin BDS: significance for CNS and biophysical studies. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25(38):8735-8745.

113. Ishida M, Yokoyama A, Shimakura K, Nagashima Y, & Shiomi K (1997) Halcurin, a polypeptide toxin from the sea anemone *Halcurias sp.*, with a structural resemblance to type 1 and 2 toxins. *Toxicon* 35(4):537-544.

114. Moran Y & Gurevitz M (2006) When positive selection of neurotoxin genes is missing - The riddle of the sea anemone *Nematostella vectensis*. *Febs Journal* 273(17):3886-3892.

115. Widmer H, Billeter M, & Wuthrich K (1989) Three-dimensional structure of the

neurotoxin ATX Ia from *Anemonia sulcata* in aqueous solution determined by nuclear magnetic resonance spectroscopy. *Proteins* 6(4):357-371.

116. Fogh RH, Kem WR, & Norton RS (1990) Solution structure of neurotoxin I from the sea anemone *Stichodactyla helianthus.* A nuclear magnetic resonance, distance geometry, and restrained molecular dynamics study. *J Biol Chem* 265(22):13016-13028.

117. Pallaghy PK, Scanlon MJ, Monks SA, & Norton RS (1995) Three-dimensional structure in solution of the polypeptide cardiac stimulant anthopleurin-A. *Biochemistry* 34(11):3782-3794.

118. Monks SA, Pallaghy PK, Scanlon MJ, & Norton RS (1995) Solution structure of the cardiostimulant polypeptide anthopleurin-B and comparison with anthopleurin-A. *Structure* 3(8):791-803.

119. Salceda E*, et al.* (2007) CgNa, a type I toxin from the giant Caribbean sea anemone *Condylactis gigantea* shows structural similarities to both type I and II toxins, as well as distinctive structural and functional properties(1). *The Biochemical journal* 406(1):67-76.

120. Catterall WA & Beress L (1978) Sea anemone toxin and scorpion toxin share a common receptor site associated with the action potential sodium ionophore. *J Biol Chem* 253(20):7393-7396.

121. Norton TR, Shibata S, Kashiwagi M, & Bentley J (1976) Isolation and characterization of the cardiotonic polypeptide anthopleurin-A from the sea anemone *Anthopleura xanthogrammica*. *J Pharm Sci* 65(9):1368-1374.

122. Cariello L*, et al.* (1989) Calitoxin, a neurotoxic peptide from the sea anemone *Calliactis parasitica*: amino acid sequence and electrophysiological properties. *Biochemistry* 28(6):2484-2489.

123. Spagnuolo A, Zanetti L, Cariello L, & Piccoli R (1994) Isolation and characterization of two genes encoding calitoxins, neurotoxic peptides from *Calliactis parasitica* (Cnidaria). *Gene* 138(1-2):187-191.

124. Torres AM & Kuchel PW (2004) The beta-defensin-fold family of polypeptides. *Toxicon* 44(6):581-588.

125. Osmakov DI*, et al.* (2013) Sea Anemone Peptide with Uncommon β-Hairpin Structure Inhibits Acid-sensing Ion Channel 3 (ASIC3) and Reveals Analgesic Activity. *Journal of Biological Chemistry* 288(32):23116-23127.

126. Zaharenko AJ*, et al.* (2008) Proteomics of the neurotoxic fraction from the sea anemone *Bunodosoma cangicum* venom: Novel peptides belonging to new classes of toxins. *Comp Biochem Physiol Part D Genomics Proteomics* 3(3):219-225.

127. Honma T*, et al.* (2008) Novel peptide toxins from the sea anemone *Stichodactyla haddoni*. *Peptides* 29(4):536-544.

128. Cassoli JS*, et al.* (2013) The proteomic profile of *Stichodactyla duerdeni* secretion reveals the presence of a novel O-linked glycopeptide. *J Proteomics* 87:89-102.

129. Logashina YA*, et al.* (2017) Peptide from Sea Anemone Metridium senile Affects Transient Receptor Potential Ankyrin-repeat 1 (TRPA1) Function and Produces

Analgesic Effect. *J Biol Chem* 292(7):2992-3004.

130. Shiomi K*, et al.* (2003) An epidermal growth factor-like toxin and two sodium channel toxins from the sea anemone *Stichodactyla gigantea*. *Toxicon* 41(2):229-236.

131. Cuypers E, Peigneur S, Debaveye S, Shiomi K, & Tytgat J (2011) TRPV1 Channel as New Target for Marine Toxins: Example of Gigantoxin I, a Sea Anemone Toxin Acting Via Modulation of the PLA_2 Pathway. *Acta Chim Slov* 58(4):735-741.

132. Orts DJ*, et al.* (2013) BcsTx3 is a founder of a novel sea anemone toxin family of potassium channel blocker. *FEBS J* 280(19):4839-4852.

133. Tudor JE, Pallaghy PK, Pennington MW, & Norton RS (1996) Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. *Nat Struct Biol* 3(4):317-320.

134. Castaneda O*, et al.* (1995) Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. *Toxicon* 33(5):603-613.

135. Chi V*, et al.* (2012) Development of a sea anemone toxin as an immunomodulator for therapy of autoimmune diseases. *Toxicon* 59(4):529-546.

136. Dauplais M*, et al.* (1997) On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *J Biol Chem* 272(7):4302-4309.

137. Gasparini S*, et al.* (1998) Delineation of the functional site of α-dendrotoxin. The functional topographies of dendrotoxins are different but share a conserved core with those of other K_V1 potassium channel-blocking toxins. *J Biol Chem* 273(39):25393-25403.

138. Kalman K*, et al.* (1998) ShK-Dap22, a potent $K_V1.3$-specific immunosuppressive polypeptide. *J Biol Chem* 273(49):32697-32707.

139. Lanigan MD*, et al.* (2002) Mutating a critical lysine in ShK toxin alters its binding configuration in the pore-vestibule region of the voltage-gated potassium channel, $K_V1.3$. *Biochemistry* 41(40):11963-11971.

140. Rashid MH & Kuyucak S (2012) Affinity and Selectivity of ShK Toxin for the K_V1 Potassium Channels from Free Energy Simulations. *J Phys Chem B* 116(16):4812-4822.

141. Upadhyay SK*, et al.* (2013) Selective $K_V1.3$ channel blocker as therapeutic for obesity and insulin resistance. *P Natl Acad Sci USA* 110(24):E2239-E2248.

142. King GF (2015) *Venoms to drugs : venom as a source for the development of human therapeutics* (Royal Society of Chemistry, Cambridge) pp xvii, 320 pages.

143. Beraud E*, et al.* (2006) Block of neural $K_V1.1$ potassium channels for neuroinflammatory disease therapy. *Annals of neurology* 60(5):586-596.

144. Antuch W, Berndt KD, Chavez MA, Delfin J, & Wuthrich K (1993) The NMR solution structure of a Kunitz-type proteinase inhibitor from the sea anemone *Stichodactyla helianthus*. *Eur J Biochem* 212(3):675-684.

145. Wunderer G, Beress L, Machleidt W, & Fritz H (1976) Broad-specificity inhibitors from sea anemones. *Methods Enzymol* 45:881-888.

146. Fritz H, Brey B, & Beress L (1972) Polyvalent isoinhibitors for trypsin, chymotrypsin, plasmin and kallikreins of sea anemones *Anemonia sulcata*, isolation, inhibitory behavior and amino acid composition. *Hoppe Seylers Z Physiol Chem* 353(1):19-30.

147. Minagawa S, Sugiyama M, Ishida M, Nagashima Y, & Shiomi K (2008) Kunitz-type protease inhibitors from acrorhagi of three species of sea anemones. *Comp Biochem Phys B* 150(2):240-245.

148. Frazao B, Vasconcelos V, & Antunes A (2012) Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: An overview. *Mar Drugs* 10(8):1812-1851.

149. Peigneur S*, et al.* (2011) A bifunctional sea anemone peptide with Kunitz type protease and potassium channel inhibiting properties. *Biochem Pharmacol* 82(1):81-90.

150. Mourao CBF & Schwartz EF (2013) Protease Inhibitors from Marine Venomous Animals and Their Counterparts in Terrestrial Venomous Animals. *Marine Drugs* 11(6):2069-2112.

151. Schweitz H*, et al.* (1995) Kalicludines and Kaliseptine — Two Different Classes of Sea-Anemone Toxins for Voltage-Sensitive K^+ Channels. *Journal of Biological Chemistry* 270(42):25121-25126.

152. Zykova TA, Vinokurov LM, Markova LF, Kozlovskaya EP, & Elyakov GB (1985) Amino-Acid-Sequence of Trypsin Inhibitor-Iv from Radianthus-Macrodactylus. *Bioorg Khim+* 11(3):293-301.

153. Sokotun IN*, et al.* (2007) A serine protease inhibitor from the anemone Radianthus macrodactylus: Isolation and physicochemical characteristics. *Russ J Bioorg Chem+* 33(4):415-422.

154. Sokotun IN*, et al.* (2006) Interaction investigation of trypsin inhibitor from sea anemone *Radianthus macrodactylus* with proteases. *Biomed Khim* 52(6):595-600.

155. Kozlov SA*, et al.* (2009) New polypeptide components from the *Heteractis crispa* sea anemone with analgesic activity. *Bioorg Khim* 35(6):789-798.

156. Andreev YA*, et al.* (2008) Analgesic compound from sea anemone *Heteractis crispa* is the first polypeptide inhibitor of vanilloid receptor 1 (TRPV1). *Journal of Biological Chemistry* 283(35):23914-23921.

157. Andreev YA, Kozlov SA, Kozlovskaya EP, & Grishin EV (2009) Analgesic effect of a polypeptide inhibitor of the TRPV1 receptor in noxious heat pain models. *Doklady. Biochemistry and biophysics* 424:46-48.

158. Cuypers E, Yanagihara A, Karlsson E, & Tytgat J (2006) Jellyfish and other cnidarian envenomations cause pain by affecting TRPV1 channels. *FEBS Lett* 580(24):5728-5732.

159. Andreev YA*, et al.* (2013) Polypeptide Modulators of TRPV1 Produce Analgesia without Hyperthermia. *Marine Drugs* 11(12):5100-5115.

160. Sunagawa S, DeSalvo MK, Voolstra CR, Reyes-Bermudez A, & Medina M (2009) Identification and gene expression analysis of a taxonomically restricted cysteine-rich protein family in reef-building corals. *Plos One* 4(3):e4865.

161. Fukuda I*, et al.* (2003) Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton. *Biochem Biophys Res Commun* 304(1):11-17.

162. Jouiaei M*, et al.* (2015) Evolution of an ancient venom: recognition of a novel family of cnidarian toxins and the common evolutionary origin of sodium and potassium neurotoxins in sea anemone. *Mol Biol Evol* 32(6):1598-1610.

163. Frazao B, Vasconcelos V, & Antunes A (2012) Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. *Mar. Drugs* 10(8):1812–1851.

164. Richet C (1903) De la thalassine toxine cristallisée et pruritogène. *C.R. Soc. Biol. Paris* 55 707–710.

165. Richet C (1905) De l'action de la congestine (virus des Actinies) sur les lapins et de ses effects anaphylactiques. *C.R. Soc. Biol. Paris* 58:109–112.

166. Nigrelli RF (1960) Biochemistry and pharmacology of compounds derived from marine organisms. *Ann. N.Y. Acad. Sci.* 90:617–949.

167. Wunderer G, Fritz H, Wachter E, & Machleidt W (1976) Amino-acid sequence of a coelenterate toxin: toxin II from A*nemonia sulcata*. *Eur J Biochem* 68(1):193–198.

168. Rodriguez AA*, et al.* (2012) 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* 34(1):26–38.

169. Honma T*, et al.* (2005) Novel peptide toxins from acrorhagi, aggressive organs of the sea anemone *Actinia equina*. *Toxicon* 46(7):768-774.

170. Macrander J, Brugler MR, & Daly M (2015) A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps. *BMC Genomics* 16:221.

171. Escoubas P, Sollod BL, & King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* 47:650–663.

172. Dutertre S*, et al.* (2013) Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* 12(2):312–329.

173. Undheim EA*, et al.* (2014) Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (centipedes). *Mol. Biol. Evol.* 31(8):2124–2148.

174. Rodriguez de la Vega RC, Corzo G, & Possani LD (2015) Scorpion venoms as a platform for drug development. *Venoms to drugs: venoms as a source for the development of human therapeutics*, ed King GF (Royal Society of Chemistry, London, UK), pp 204–220.

175. Walker AA*, et al.* (2017) Melt with this kiss: paralysing and liquefying venom of the assassin bug *Pristhesancus plagipennis* (Hemiptera: Reduviidae). *Mol. Cell. Proteomics*:in press (doi:10.1074/mcp.M1116.063321).

176. Martins RD*, et al.* (2009) Purification and characterization of the biological effects of phospholipase A_2 from sea anemone *Bunodosoma caissarum*. *Toxicon* 54(4):413–420.

177. Macrander J, Broe M, & Daly M (2015) Multi-copy venom genes hidden in de novo transcriptome assemblies, a cautionary tale with the snakelocks sea anemone

Anemonia sulcata (Pennant, 1977). *Toxicon* 108:184–188.

178. Cassoli JS*, et al.* (2013) The proteomic profile of *Stichodactyla duerdeni* secretion reveals the presence of a novel O-linked glycopeptide. *J. Proteomics* 87:89–102.

179. Malpezzi EL, de Freitas JC, Muramoto K, & Kamiya H (1993) Characterization of peptides in sea anemone venom collected by a novel procedure. *Toxicon* 31(7):853–864.

180. Hale JE, Butler JP, Gelfanova V, You JS, & Knierman MD (2004) A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Anal Biochem* 333(1):174–181.

181. Vizcaino JA*, et al.* (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44(D1):D447-456.

182. Vizcaino JA*, et al.* (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* 41(Database issue):D1063-1069.

183. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

184. Grabherr MG*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* 29(7):644–652.

185. Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

186. Afgan E*, et al.* (2015) Genomics Virtual Laboratory: a practical bioinformatics workbench for the cloud. *PloS One* 10(10):e0140829.

187. Robinson JT*, et al.* (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24-26.

188. Thorvaldsdottir H, Robinson JT, & Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178-192.

189. Cock PJA, Gruning BA, Paszkiewicz K, & Pritchard L (2013) Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* 1:e167.

190. Jungo F, Bougueleret L, Xenarios I, & Poux S (2012) The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. *Toxicon* 60(4):551–557.

191. Fu LM, Niu BF, Zhu ZW, Wu ST, & Li WZ (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152.

192. Petersen TN, Brunak S, von Heijne G, & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.

193. Mitchell A*, et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43:D213–D221.

194. Conesa A*, et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.

195. Balasubramanian PG*, et al.* (2012) Proteome of Hydra nematocyst. *J. Biol. Chem.* 287(13):9672–9781.

196. Graudins A*, et al.* (2012) Cloning and activity of a novel α-latrotoxin from red-back spider venom. *Biochem. Pharmacol.* 83(1):170–183.

197. Ushkaryov YA, Volynski KE, & Ashton AC (2004) The multiple actions of black widow spider toxins and their selective use in neurosecretion studies. *Toxicon* 43(5):527–542.

198. Garb JE & Hayashi CY (2013) Molecular evolution of α-latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol. Biol. Evol.* 30(5):999–1014.

199. Whittington CM*, et al.* (2010) Novel venom gene discovery in the platypus. *Genome Biol* 11(9):R95.

200. de Oliveira NG*, et al.* (2016) Venom gland transcriptome analyses of two freshwater stingrays (Myliobatiformes: Potamotrygonidae) from Brazil. *Sci. Rep.* 6:21935.

201. Rodriguez de la Vega RC & Giraud T (2016) Intragenome diversity of gene families encoding toxin-like proteins in venomous animals. *Integr. Comp. Biol.* 56(5):938–949.

202. Matsushita M (2010) Ficolins: complement-activating lectins involved in innate immunity. *J. Innate Immun.* 2(1):24–32.

203. OmPraba G*, et al.* (2010) Identification of a novel family of snake venom proteins Veficolins from *Cerberus rynchops* using a venom gland transcriptomics and proteomics approach. *J. Proteome Res.* 9(4):1882–1893.

204. Fry BG*, et al.* (2010) Functional and structural diversification of the Anguimorpha lizard venom system. *Mol. Cell. Proteomics* 9(11):2369–2390.

205. Davidson CJ, Tuddenham EG, & McVey JH (2003) 450 million years of hemostasis. *J. Thromb. Haemost.* 1(7):1487–1494.

206. Rao VS, Swarup S, & Kini RM (2003) The nonenzymatic subunit of pseutarin C, a prothrombin activator from eastern brown snake (*Pseudonaja textilis*) venom, shows structural similarity to mammalian coagulation factor V. *Blood* 102(4):1347–1354.

207. Suarez-Carmona M, Hubert P, Delvenne P, & Herfs M (2015) Defensins: "simple" antimicrobial peptides or broad-spectrum molecules? *Cytokine Growth Factor Rev.* 26(3):361–370.

208. Chagot B*, et al.* (2005) Solution structure of APETx2, a specific peptide inhibitor of ASIC3 proton-gated channels. *Protein Sci.* 14(8):2003–2010.

209. Undheim EA, Mobli M, & King GF (2016) Toxin structures as evolutionary tools: using conserved 3D folds to study the evolution of rapidly evolving peptides. *Bioessays* 38(6):539–548.

210. Saez NJ*, et al.* (2010) Spider-venom peptides as therapeutics. *Toxins* 2(12):2851–2871.

211. Herzig V & King GF (2015) The cystine knot is responsible for the exceptional stability of the insecticidal spider toxin ω-hexatoxin-Hv1a. *Toxins* 7(10):4366–4380.

212. Sollod BL*, et al.* (2005) Were arachnids the first to use combinatorial peptide libraries? *Peptides* 26:131–139.

213. Robinson SD*, et al.* (2014) Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS One* 9(2):e87648.

214. Mourao CB & Schwartz EF (2013) Protease inhibitors from marine venomous animals and their counterparts in terrestrial venomous animals. *Mar Drugs* 11(6):2069-2112.

215. Fry BG*, et al.* (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* 10:483–511.

216. Garcia-Fernandez R*, et al.* (2016) The Kunitz-Type Protein ShPI-1 Inhibits Serine Proteases and Voltage-Gated Potassium Channels. *Toxins (Basel)* 8(4):110.

217. Honma T & Shiomi K (2006) Peptide toxins in sea anemones: structural and functional aspects. *Mar. Biotech.* 8(1):1–10.

218. Archer J, Whiteley G, Casewell NR, Harrison RA, & Wagstaff SC (2014) VTBuilder: a tool for the assembly of multi isoform transcriptomes. *BMC Bioinformatics* 15:389.

219. Sunagar K, Morgenstern D, Reitzel AM, & Moran Y (2016) Ecological venomics: how genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. *J. Proteomics* 135:62–72.

220. Sunagar K & Moran Y (2015) The rise and fall of an evolutionary innovation: contrasting strategies of venom evolution in ancient and young animals. *PLOS Genet.* 11(10):e1005596.

221. Morgenstern D*, et al.* (2011) The tale of a resting gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon* 57(5):695–703.

222. Rodriguez de la Vega RC & Giraud T (2016) Intragenome diversity of gene families encoding toxin-like proteins in venomous animals. *Integr. Comp. Biol.* 56:938–949.

223. Ozbek S (2011) The cnidarian nematocyst: a miniature extracellular matrix within a secretory vesicle. *Protoplasma* 248(4):635-640.

224. Brinkman DL*, et al.* (2012) Venom proteome of the box jellyfish *Chironex fleckeri*. *PLoS One* 7(12):e47866.

225. Brinkman DL*, et al.* (2015) Transcriptome and venom proteome of the box jellyfish *Chironex fleckeri*. *BMC Genomics* 16:407.

226. Anderluh G, Podlesek Z, & Macek P (2000) A common motif in proparts of Cnidarian toxins and nematocyst collagens and its putative role. *Biochim Biophys Acta* 1476(2):372-376.

227. Van Iten H, Marques AC, Leme JdM, Pacheco MLAF, & Simões MG (2014) Origin and early diversification of the phylum Cnidaria Verrill: major developments in the analysis of the taxon's Proterozoic–Cambrian history. *Palaeontology* 57(4):677-690.

228. Remigante A*, et al.* (2018) Impact of Scyphozoan Venoms on Human Health and Current First Aid Options for Stings. *Toxins (Basel)* 10(4).

229. Pennington MW, Czerwinski A, & Norton RS (2017) Peptide therapeutics from venom: Current status and potential. *Bioorg Med Chem*.

230. Chapman JA*, et al.* (2010) The dynamic genome of Hydra. *Nature* 464(7288):592-596.

231. Shinzato C*, et al.* (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476(7360):320-323.

232. Park E*, et al.* (2012) Estimation of divergence times in cnidarian evolution based on

mitochondrial protein-coding genes and the fossil record. *Mol Phylogenet Evol* 62(1):329-345.

233. Menon LR, McIlroy D, & Brasier MD (2013) Evidence for Cnidaria-like behavior in ca. 560 Ma Ediacaran Aspidella. *Geology* 41(8):895-898.

234. Erwin DH*, et al.* (2011) The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. *Science* 334(6059):1091-1097.

235. Jouiaei M*, et al.* (2015) Ancient venom systems: a review on Cnidaria toxins. *Toxins* 7(6):2251–2271.

236. Prentis PJ, Pavasovic A, & Norton RS (2018) Sea Anemones: Quiet Achievers in the Field of Peptide Toxins. *Toxins (Basel)* 10(1).

237. Columbus-Shenkar YY*, et al.* (2018) Dynamics of venom composition across a complex life cycle. *Elife* 7.

238. Ollerton J, McCollin D, Fautin DG, & Allen GR (2007) Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proc Biol Sci* 274(1609):591-598.

239. Nedosyko AM, Young JE, Edwards JW, & Burke da Silva K (2014) Searching for a toxic key to unlock the mystery of anemonefish and anemone symbiosis. *PLoS One* 9(5):e98449.

240. International Society for Reef Studies. & SpringerLink (Online service) (Coral reefs. (Springer, Berlin ; Heidelberg).

241. Lieske E & Myers R (2001) *Coral reef fishes : Indo-Pacific and Caribbean* (HarperCollins, London) Rev. Ed p 400 p.

242. Undheim EA, Mobli M, & King GF (2016) Toxin structures as evolutionary tools: Using conserved 3D folds to study the evolution of rapidly evolving peptides. *Bioessays* 38(6):539-548.

243. Lee CW*, et al.* (2004) Molecular basis of the high-affinity activation of type 1 ryanodine receptors by imperatoxin A. *Biochem J* 377(Pt 2):385-394.

244. Bernard C*, et al.* (2004) Solution structure of ADO1, a toxin extracted from the saliva of the assassin bug, *Agriosphodrus dohrni*. *Proteins* 54(2):195-205.

245. Daly NL*, et al.* (2004) Structures of muO-conotoxins from *Conus marmoreus*. Inhibitors of tetrodotoxin (TTX)-sensitive and TTX-resistant sodium channels in mammalian sensory neurons. *J Biol Chem* 279(24):25774-25782.

246. Vink S, Daly NL, Steen N, Craik DJ, & Alewood PF (2014) Holocyclotoxin-1, a cystine knot toxin from *Ixodes holocyclus*. *Toxicon* 90:308-317.

247. Torres AF*, et al.* (2014) Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadriceps*: insights into the polypeptide toxin arsenal of hymenopterans. *PLoS One* 9(1):e87556.

248. von Reumont BM*, et al.* (2014) The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Mol Biol Evol* 31(1):48-58.

249. Undheim EA*, et al.* (2014) Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (centipedes). *Mol Biol Evol* 31(8):2124-2148.

250. Peigneur S*, et al.* (2011) A bifunctional sea anemone peptide with Kunitz type protease and potassium channel inhibiting properties. *Biochem Pharmacol* 82(1):81-90.

251. Liao Q*, et al.* (2018) Novel Kunitz-like Peptides Discovered in the Zoanthid *Palythoa caribaeorum* through Transcriptome Sequencing. *J Proteome Res* 17(2):891-902.

252. Helm RR, Siebert S, Tulin S, Smith J, & Dunn CW (2013) Characterization of differential transcript abundance through time during *Nematostella vectensis* development. *BMC Genomics* 14:266.

253. Stefanik DJ*, et al.* (2014) Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, *Edwardsiella lineata*, a parasitic cnidarian. *BMC Genomics* 15:71.

254. Lehnert EM, Burriesci MS, & Pringle JR (2012) Developing the anemone Aiptasia as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic *A. pallida*. *BMC Genomics* 13:271.

255. Macrander J & Daly M (2016) Evolution of the Cytolytic Pore-Forming Proteins (Actinoporins) in Sea Anemones. *Toxins (Basel)* 8(12).

256. Hwang JS*, et al.* (2007) The evolutionary emergence of cell type-specific genes inferred from the gene expression analysis of Hydra. *Proc Natl Acad Sci U S A* 104(37):14735-14740.

257. Madio B, Undheim EAB, & King GF (2017) Revisiting venom of the sea anemone *Stichodactyla haddoni*: Omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus. *Journal of Proteomics*.

258. Grabherr MG*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.

259. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357-359.

260. Robinson JT*, et al.* (2011) Integrative Genomics Viewer. *Nat Biotechnol* 29(1):24-26.

261. Thorvaldsdóttir H, Robinson JT, & Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14(2):178-192.

262. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4):772–780.

263. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular biology and evolution* 32(1):268-274.

264. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* 14(6):587-589.

265. Minh BQ, Nguyen MAT, & von Haeseler A (2013) Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular biology and evolution* 30(5):1188-1195.

266. Ciccarelli FD*, et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311(5765):1283-1287.

267. Fry BG*, et al.* (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genom. Human Genet.* 10:483–511.

268. Jenner R & Undheim E (2017) *Venom: the secrets of nature's deadliest weapon* (Natural History Museum, London, UK).

269. King GF (2015) *Venoms to drugs: venom as a source for the development of human therapeutics* (Royal Society of Chemistry, London, UK).

270. Casewell NR, Wuster W, Vonk FJ, Harrison RA, & Fry BG (2013) Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology and Evolution* 28(4):219-229.

271. Mikov AN & Kozlov SA (2015) Structural features of cysteine-rich polypeptides from sea anemone venoms. *Russian J. Bioorg. Chem.* 41(5):455–466.

272. Logashina YA*, et al.* (2017) New disulfide-stabilized fold provides sea anemone peptide to exhibit both antimicrobial and TRPA1 potentiating properties. *Toxins* 9(5):154.

273. Madio B, Undheim EAB, & King GF (2017) Revisiting venom of the sea anemone *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus. *J. Proteomics* 166:83–92.

274. Tarcha EJ*, et al.* (2017) Safety and pharmacodynamics of dalazatide, a K_V1.3 channel inhibitor, in the treatment of plaque psoriasis: a randomized phase 1b trial. *PLoS One* 12(7):e0180762.

275. Moran Y*, et al.* (2008) Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the *Nematostella vectensis* genome. *Mol. Biol. Evol.* 25(4):737–747.

276. Davis IW*, et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:375–383.

277. Panteleev PV, Balandin SV, Ivanov VT, & Ovchinnikova TV (2017) A therapeutic potential of animal β-hairpin antimicrobial peptides. *Curr. Med. Chem.* 24(17):1724–1746.

278. Chagot B*, et al.* (2005) An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*. *Biochem. J.* 388(Pt 1):263–271.

279. Smith JJ*, et al.* (2011) Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif. *Proc. Natl. Acad. Sci. USA* 108(26):10478–10483.

280. Yang S*, et al.* (2015) A pain-inducing centipede toxin targets the heat activation machinery of nociceptor TRPV1. *Nat. Commun.* 6:8297.

281. Edwards IA*, et al.* (2016) Contribution of amphipathicity and hydrophobicity to the antimicrobial activity and cytotoxicity of β-hairpin peptides. *ACS Infect. Dis.* 2(6):442–

450.

282. Henriques ST*, et al.* (2017) Redesigned spider peptide with improved antimicrobial and anticancer properties. *ACS Chem. Biol.* 12(9):2324–2334.

283. Basulto A*, et al.* (2006) Immunohistochemical targeting of sea anemone cytolysins on tentacles, mesenteric filaments and isolated nematocysts of *Stichodactyla helianthus*. *J. Exp. Zool. Part A, Comp. Exp. Biol.* 305(3):253–258.

284. Beckmann A & Ozbek S (2012) The nematocyst: a molecular map of the cnidarian stinging organelle. *Int. J. Dev. Biol.* 56(6-8):577–582.

285. Undheim EAB*, et al.* (2014) Multifunctional warheads: diversification of the toxin arsenal of centipedes via novel multidomain transcripts. *J. Proteomics* 102:1–10.

286. Undheim EAB*, et al.* (2015) Production and packaging of a biological arsenal: evolution of centipede venoms under morphological constraint. *Proc. Natl. Acad. Sci. USA* 112(13):4026–4031.

287. Mitchell ML*, et al.* (2017) The use of imaging mass spectrometry to study peptide toxin distribution in Australian sea anemones. *Austr. J. Chem.* 70(11):1235–1237.

288. Chintiroglou C & Koukouras A (1992) The feeding habits of three Mediterranean sea anemone species, *Anemonia viridis* (Forskål), *Actinia equina* (Linnaeus) and *Cereus pedunculatus* (Pennant). *Helgoländer Meeresuntersuchungen* 46(1):53-68.

289. Kruger LM & Griffiths CL (1996) Sources of nutrition in intertidal sea anemones from the south-western Cape, South Africa. *S. Afr. J. Zool.* 31(3):110–119.

290. Ottaway JR (1978) Population ecology of the intertidal anemone *Actinia tenebrosa*: pedal locomotion and intraspecific aggression. *Austr. J. Mar. Freshwater Res.* 29(6):787–802.

291. Robson SA & King GF (2006) Domain architecture and structure of the bacterial cell division protein DivIB. *Proc. Natl. Acad. Sci. USA* 103(17):6700–6705.

292. Lummis SC*, et al.* (2005) *Cis-trans* isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* 438(7065):248–252.

293. Andreotti AH (2006) Opening the pore hinges on proline. *Nat. Chem. Biol.* 2(1):13–14.

294. Hille B (2001) *Ion channels of excitable membranes* (Sinauer, Sunderland, MA, USA) 3rd Ed.

295. Bende NS*, et al.* (2014) A distinct sodium channel voltage-sensor locus determines insect selectivity of the spider toxin Dc1a. *Nat. Commun.* 5:4350.

296. Undheim EAB, Mobli M, & King GF (2016) Toxin structures as evolutionary tools: Using conserved 3D folds to study the evolution of rapidly evolving peptides. *BioEssays : news and reviews in molecular, cellular and developmental biology* 38(6):539-548.

297. Brown DD, Wensink PC, & Jordan E (1972) A comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.* 63(1):57–73.

298. Schüler A & Bornberg-Bauer E (2016) Evolution of protein domain repeats in metazoa. *Mol. Biol. Evol.* 33(12):3170–3182.

299. Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, & Rokyta DR (2017) Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics* 206(3):1569-1580.

300. Fukuyama Y, Iwamoto S, & Tanaka K (2006) Rapid sequencing and disulfide mapping of peptides containing disulfide bonds by using 1,5-diaminonaphthalene as a reductive matrix. *Journal of Mass Spectrometry* 41(2):191-201.

301. Caprioli RM, Farmer TB, & Gile J (1997) Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Analytical Chemistry* 69(23):4751-4760.

302. Yarnold JE*, et al.* (2012) High resolution spatial mapping of brominated pyrrole-2-aminoimidazole alkaloids distributions in the marine sponge *Stylissa flabellata* via MALDI-mass spectrometry imaging. *Molecular Biosystematics* 8(9):2249-2259.

303. Undheim EAB*, et al.* (2014) Multifunctional warheads: diversification of the toxin arsenal of centipedes via novel multidomain transcripts. *Journal of Proteomics* 102:1-10.

304. Hale JE, Butler JP, Gelfanova V, You JS, & Knierman MD (2004) A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Anal Biochem* 333(1):174–181.

305. Cock PJ, Gruning BA, Paszkiewicz K, & Pritchard L (2013) Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ* 1:e167.

306. Camacho C*, et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421.

307. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32(1):268–274.

308. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Meth.* 14(6):587–589.

309. Minh BQ, Nguyen MAT, & von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30(5):1188–1195.

310. Vranken WF*, et al.* (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59(4):687-696.

311. Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods in molecular biology* 278:353-378.

312. Tytgat J, Debont T, Carmeliet E, & Daenens P (1995) The α-dendrotoxin footprint on a mammalian potassium channel. *J Biol Chem* 270(42):24776-24781.

313. Henriques ST*, et al.* (2012) Phosphatidylethanolamine binding is a conserved feature of cyclotide-membrane interactions. *Journal of Biological Chemistry* 287(40):33629-33643.

314. Henriques ST*, et al.* (2011) Decoding the membrane activity of the cyclotide kalata B1: the importance of phosphatidylethanolamine phospholipids and lipid organization on

hemolytic and anti-HIV activities. *Journal of Biological Chemistry* 286(27):24231-24241.

315. Henriques ST, Pattenden LK, Aguilar MI, & Castanho MA (2008) PrP(106-126) does not interact with membranes under physiological conditions. *Biophysical Journal* 95(4):1877-1889.

316. Huang YH, Colgrave ML, Clark RJ, Kotze AC, & Craik DJ (2010) Lysine-scanning mutagenesis reveals an amendable face of the cyclotide kalata B1 for the optimization of nematocidal activity. *Journal of Biological Chemistry* 285(14):10797-10805.

317. Torcato IM*, et al.* (2013) Design and characterization of novel antimicrobial peptides, R-BP100 and RW-BP100, with activity against Gram-negative and Gram-positive bacteria. *Biochimica et Biophysica Acta* 1828(3):944-955.

318. Oliveira JS*, et al.* (2006) BcIV, a new paralyzing peptide obtained from the venom of the sea anemone *Bunodosoma caissarum*. A comparison with the Na+ channel toxin BcIII. *Biochimica et Biophysica Acta* 1764(10):1592-1600.

Appendices

Collection Permit

QUEENSLAND GOVERNMENT Department of National Parks, Recreation, Sport and Racing

Permit

Marine Parks Act 2004 PERMIT NO. **QS2014/MAN259**

TERM OF PERMISSION:

20 May 2014 to 31 December 2016

THIS PERMISSION IS GRANTED TO:

Permit holder: Bruno Madio

Place of Business: Institute of Molecular Bioscience
University of Queensland
306 Carmody Road
ST LUCIA QLD 4072

for use and entry of **Moreton Bay Marine Park**

in accordance with the details as stated in Part A, and subject to the conditions stated in Part B.

Signed: ..... Date: 16/5/14

Damian McCarthy
Delegate of the Chief Executive,
Department of National Parks, Recreation, Sport and Racing

Part A:
LOCATION TO WHICH THIS PERMISSION APPLIES:

Zone(s): CPZ07 – South Passage – Boat Rock
HPZ12 – Flinders Beach
HPZ02 – Moreton Island – Broadwater
GUZ08 - Dunwich

PURPOSE OF USE OR ENTRY AUTHORISED BY THIS PERMISSION:

Conduct of research involving the collection of sea anemones

Part B:
CONDITIONS OF PERMISSION:

1. Location
Activities described in the 'purpose of use and entry' of this permit must only take place in the locations described in 'Part A' of this permit.

2. Disturbance to adjacent areas
The activity must be conducted in a manner that causes minimal disturbance to the marine ecosystem (e.g. adjacent mangroves, corals and the substrate).

3. Conduct of research
The permit holder must ensure that:-

- All research is conducted in accordance with marine park permit application dated 1 April 2014 and information provided by email on 15 May 2014.
- No tools are used to collect specimens. All sea anemones are to be collected by hand only.
- A maximum of 15 specimens per species are collected.
- Any specimens to be returned to the marine park have not been exposed to any contaminants, medications or antibiotics while held in aquaria at the University of Queensland's Institute of Molecular Bioscience.

4. Report
The following reports must be submitted:

(a) Within 20 working days of fieldwork being completed, a catch summary that includes the following information:
  (i) the identification of each species collected
  (ii) the date and location of where specimens were collected
  (iii) the numbers of each specimens kept for further analysis; and
  (iv) the numbers of each specimens released to each site.

(b) At the end of the research, a summary of the research findings.

(c) The returns must be lodged with the Administration Officer, Tourism and Recreation Access, Department of National Parks, Recreation, Sport and Racing at:

- PO Box 15187 CITY EAST Qld 4002; or
- e-mail to southernpermitsteam@nprsr.qld.gov.au

5. Indemnity

(a) The permit holder:

(i) indemnifies; and

(ii) releases and discharges

the State of Queensland ("the State") from and against all actions, proceedings, claims, demands, costs, losses, damages and expenses which may be brought against or made upon the State, or which the State may pay, sustain or be put to by reason of, or in consequence of, or in connection with the occupation and use of the marine park by the permit holder except to the extent of any negligent act of the State, its servants or agents.

(b) The permit holder must notify the Chief Executive in writing of any death, injury, loss or damage immediately upon becoming aware of such death, injury, loss or damage.

6. Compliance with laws

The permit holder must at its own expense comply with all Acts of Parliament, local laws, regulations or rules for the time being in force which apply to the marine park and/or the permit holder's use of the marine park.

7. No warranty

(a) The State does not warrant that the marine park is free from defect or that it is safe or suitable for the permitted purpose.

(b) The permit holder uses the marine park entirely at its own risk and acknowledges that it has checked the area to ensure that it is suitable for the permitted purpose.

8. Public liability insurance

(a) For the term of this permit, the permit holder must hold the following insurances:

(i) a public liability insurance policy for not less than $20 million arising from any one event in respect of the death of, or injury to persons, or loss or damage to property; and

(ii) insurance under the *Workers Compensation and Rehabilitation Act 2003* to cover workers, eligible persons, self employed contractors, directors, trustees and partners.

(b) The insurance policies required under this condition must be held with an insurer that is registered with the Australian Prudential Regulation Authority and has an S&P rating of no less than A-.

(c) The insurance policies in place to cover the insurable interests pursuant to this condition must cover all invitees, employees, contractors, agents, members or clients of the permit holder and name the State as an interested party.

(d) Before the permit holder undertakes any activities in the marine park, all insurances required by this condition must be in place.

(e) In any circumstance where the insurances required under this Permit are cancelled, altered or expire before the expiry date of this permit, the permit holder must cease all activities in the marine park until such time as alternative insurance policies that comply with the requirements of this condition have been obtained.

(f) From time to time (but no more than once annually) the State, acting reasonably, may issue a notice to the permit holder increasing the minimum insurance cover required by this condition.

(g) The permit holder must provide copies of certificates of currency or the insurance policies documents required under this condition when requested by Chief Executive for audit or compliance checks.

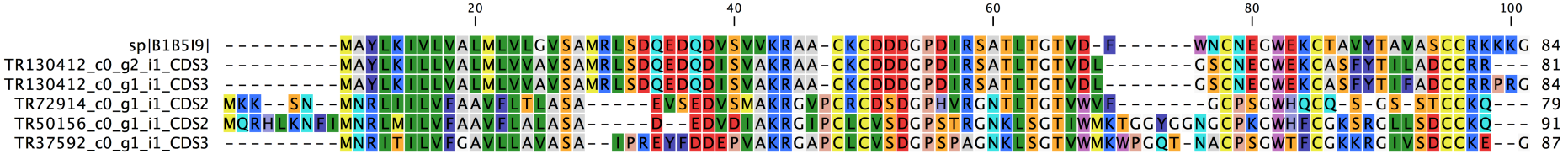
9. Breach and termination

This Permit will be terminated immediately upon the permit holder failing to comply with the provisions of the insurance and indemnity conditions.

Appendices of Chapter 2

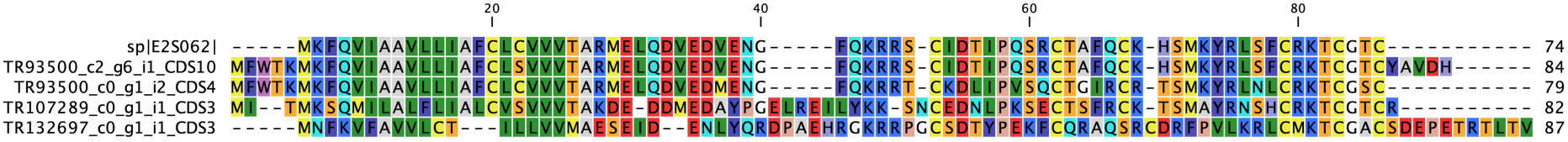
Appendix I Sequence alignment with toxins reported by Honma et al. (2008).

Delta-stichotoxin-Shd3a - B1B5I9 (NA24_STIHA)

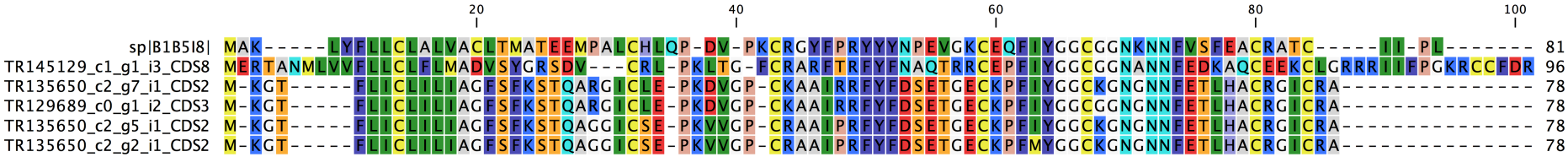


Omega-stichotoxin-Shd4a - B1B5J0 (SHTX5_STIHA)

Kappa-stichotoxin-Shd5a - E2S062 (K1A_STIHA)



KappaPI-stichotoxin-Shd2a - B1B5I8 (VKT3_STIHA)



Kappa-stichotoxin-Shd1a/kappa-stichotoxin-Shd1b - P0C7W7 (TX9A_STIHA)

Appendix II Summary of the 33 toxins identified via proteomic analysis of S. haddoni milked venom that showed significant homology to entries in the UniProt toxin database.

Transcript	Top BLAST hit	e-value	Type	Organism	Family/Scafold	Size
TR138219_c4_g1_i1_CDS10|length	sp|R4ZCU1|TX9A_URTGR,sp|R4ZCU1|TX9A_URTGR,sp|R4ZCU1|TX9A_URTGR	2e-10,1e-08,3e-10	ASIC	Sea anemone	Boundless beta-hairpin	Peptide
TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	2.00E-56	Nav type2	Sea anemone	Defensin	Peptide
TR130412_c0_g2_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	4.00E-54	Nav type2	Sea anemone	Defensin	Peptide
TR134234_c17_g1_i1_CDS4|length	sp|P0DMX5|BDS2C_ANTEL	4.00E-08	Kv type3	Sea anemone	Defensin	Peptide
TR50156_c0_g1_i1_CDS2|length	sp|Q9NJQ2|NA11_ACTEQ	6.00E-57	Nav type1	Sea anemone	Defensin	Peptide
TR72914_c0_g1_i1_CDS2|length	sp|Q76CA3|NA1G2_STIGI,sp|Q76CA3|NA1G2_STIGI	6e-50,6e-50	Nav type1	Sea anemone	Defensin	Peptide
TR72914_c0_g2_i1_CDS2|length	sp|Q76CA3|NA1G2_STIGI,sp|Q76CA3|NA1G2_STIGI	6e-50,6e-50	Nav type2	Sea anemone	Defensin	Peptide
TR18604_c0_g1_i1_CDS1|length	sp|Q76CA0|NA2G3_STIGI	2.00E-56	Nav type2	Sea anemone	Defensin	Peptide
TR69080_c0_g1_i1_CDS11|length	sp|B1B5J0|SHTX5_STIHA	3.00E-59	EGF-like sea anemone	Sea anemone	EGF-like	Peptide
TR71631_c0_g6_i1_CDS1|length	sp|B1P1G9|JZ20A_CHIGU	6.00E-05	ICK - Unknow	Spider	ICK	Peptide
TR71631_c0_g6_i2_CDS2|length	sp|B1P1G9|JZ20A_CHIGU	5.00E-05	ICK - Unknow	Spider	ICK	Peptide
TR129689_c0_g1_i2_CDS3|length	sp|B2G331|VKT1_HETCR	1.00E-42	TRPV1	Sea anemone	Kunitz-type	Peptide
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	9.00E-43	TRPV1	Sea anemone	Kunitz-type	Peptide
TR135650_c2_g5_i1_CDS2|length	sp|B2G331|VKT1_HETCR	4.00E-43	TRPV1	Sea anemone	Kunitz-type	Peptide
TR135650_c2_g6_i1_CDS2|length	sp|B2G331|VKT1_HETCR	1.00E-42	TRPV2	Sea anemone	Kunitz-type	Peptide
TR135650_c2_g7_i1_CDS2|length	sp|B2G331|VKT1_HETCR	4.00E-43	TRPV1	Sea anemone	Kunitz-type	Peptide
TR135650_c2_g8_i1_CDS2|length	sp|B2G331|VKT1_HETCR	8.00E-36	TRPV1	Sea anemone	Kunitz-type	Peptide
TR145129_c1_g1_i3_CDS8|length	sp|P0DN15|VKTB_ANEVI	5.00E-35	Kv type2	Sea anemone	Kunitz-type	Peptide
TR130412_c6_g8_i3_CDS2|length	sp|P0C7W7|TX9A_STIHA,sp|P0C7W7|TX9A_STIHA	3e-04,3e-04	Kv type4	Sea anemone	Kv type4	Peptide
TR138219_c4_g2_i1_CDS10|length	sp|P0C7W7|TX9A_STIHA	4.00E-14	Kv type4	Sea anemone	Kv type4	Peptide

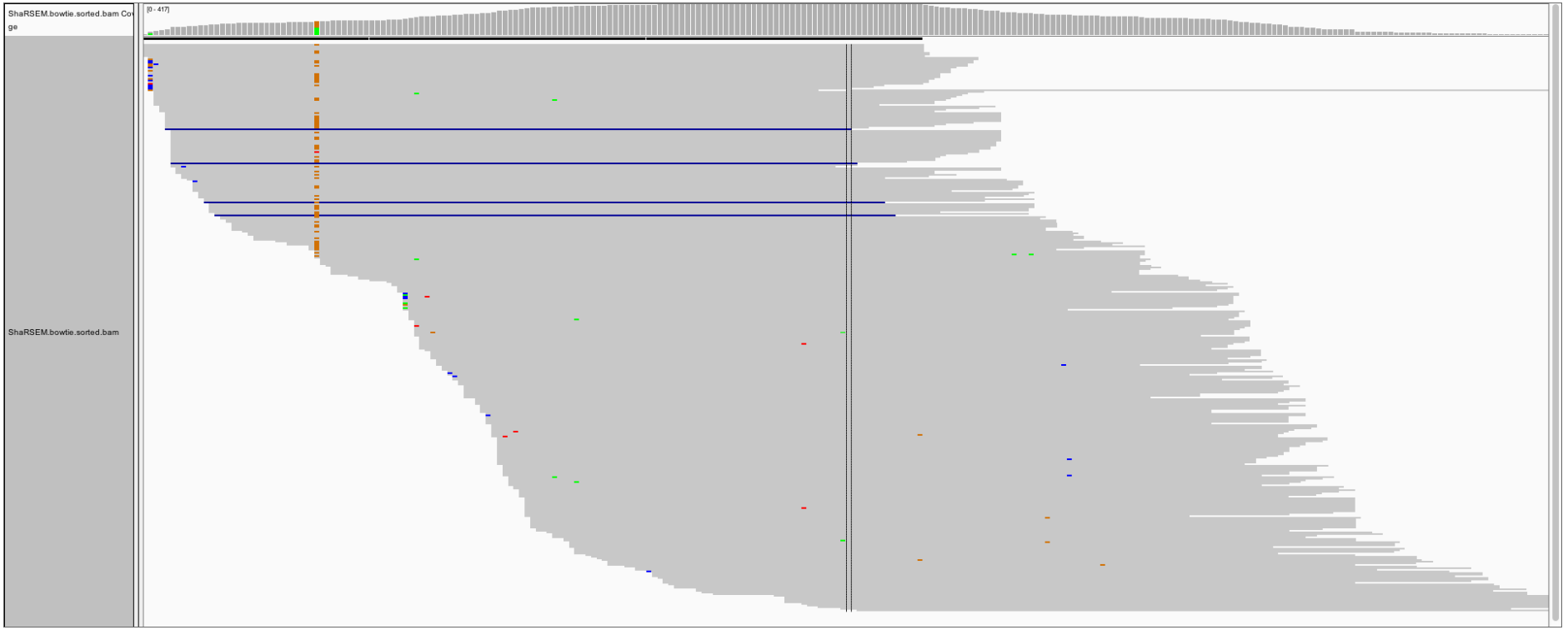
TR72284_c0_g1_i1_CDS4|length	sp|A7RMN1|KV51_NEMVE	1.00E-19	Kv Type5	Sea anemone	Kv Type5	Peptide
TR107289_c0_g1_i1_CDS3|length	sp|Q0EAE5|K1A_ANEER	1.00E-23	Kv type1	Sea anemone	ShKT domain	Peptide
TR93500_c0_g1_i2_CDS4|length	sp|O16846|K1A_HETMG	1.00E-46	Kv type1	Sea anemone	ShKT domain	Peptide
TR93500_c2_g3_i1_CDS4|length	sp|O16846|K1A_HETMG	1.00E-42	Kv type1	Sea anemone	ShKT domain	Peptide
TR96729_c4_g1_i3_CDS26|length	sp|Q9TWG1|K1B_ANESU	3.00E-10	Kv type1	Sea anemone	ShKT domain	Peptide
TR110527_c0_g1_i2_CDS3|length	sp|A9YME1|VA5_MICHY	1.00E-16	Venom allergen 5	Wasp	CAP	Protein
TR104378_c2_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	1.00E-143	Cytolysin type2	Sea anemone	Cytolysin	Protein
TR104378_c2_g2_i2_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	4.00E-139	Cytolysin type2	Sea anemone	Cytolysin	Protein
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	5.00E-06	Cytolysin type2	Sea anemone	Cytolysin	Protein
TR65582_c0_g1_i4_CDS2|length	sp|Q9U6X1|ACTP3_HETMG	8.00E-116	Cytolysin type2	Sea anemone	Cytolysin	Protein
TR65582_c0_g1_i5_CDS2|length	sp|Q9U6X1|ACTP3_HETMG	8.00E-116	Cytolysin type3	Sea anemone	Cytolysin	Protein
TR111298_c2_g1_i1_CDS2|length	sp|Q58L93|FAXD_PSEPO	1.00E-34	Prothrombin activator	Snake	S1 peptidase	Protein
TR111298_c2_g1_i2_CDS2|length	sp|Q58L93|FAXD_PSEPO	1.00E-34	Prothrombin activator	Snake	S1 peptidase	Protein

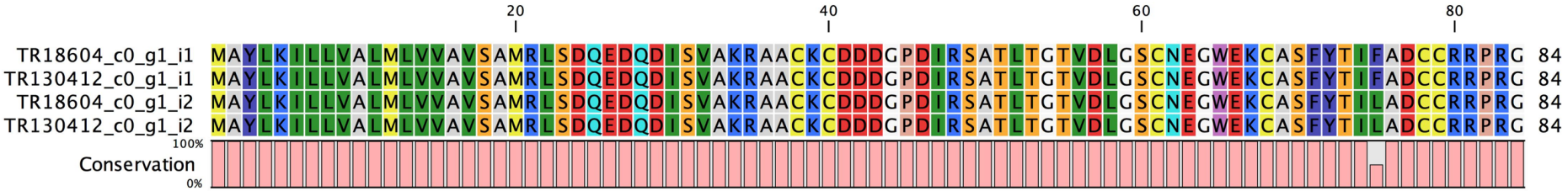
Appendix III Hidden polymorphisms

Accessions	Top BLAST hit	Type	Family/Scafold	Unused	Total	%Cov	%Cov(50)	%Cov(95)	Contrib	Conf	Sequence	Modifications	Cleavages	dMass	Prec MW	Prec m/z	Theor MW	Theor m/z	Theor z	Sc	Spectrum	Time	PrecursorSignal	PrecursorElution
TR111298_c2_g1_i2_CDS2|length; TR111298_c2_g1_i1_CDS2|length	sp|Q58L93|FAXD_PSEPO	Prothrombin activator	S1 peptidase	14.35	14.35	40.29	35.16	35.16	0	99.00	HDVALLQLAQPVTLTK	Gln->Lys@10		-0.01	1746.03	874.02	1746.04	874.03	2	21	1.1.1.1894.3	7.0831	5460	7.0635
TR130412_c6_g8_i2_CDS3|length; TR130412_c6_g8_i1_CDS2|length	sp|P0C7W7|TX9A_STIHA,sp|P0C7W7|TX9A_STIHA	Kv type4	Kv type4	15.05	15.05	44.12	42.65	42.65	0.35	99.00	AIIDRNSCRGKCNRMDHLGKCRK	Ile->Asn@2	missed R-N@5; missed R-G@9; missed K-C@11; missed R-M@14; missed K-C@20; missed R-K@22	0.03	2806.40	562.29	2806.37	562.28	5	15	1.1.1.1726.16	14.4415	141369	14.3726
TR130412_c6_g8_i3_CDS2|length	sp|P0C7W7|TX9A_STIHA,sp|P0C7W7|TX9A_STIHA	Kv type4	Kv type4	8.15	8.15	40.91	40.91	40.91	0	99.00	KCYKKDANGVCRKVFGCEP	Val->Thr@10; Glu->Gln@18	missed K-C@1; missed K-K@4; missed R-K@12; missed K-V@13	-0.01	2277.09	570.28	2277.11	570.28	4	17	1.1.1.1740.4	15.7306	128548	15.6154
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0.15	99.00	GICSEPKVVGPCRAAIPR	Arg->Lys@13 TR135650_c2_g2_i3	cleaved G-G@N-term; missed K-V@7; missed R-A@13	0.00	1912.03	638.35	1912.03	638.35	3	15	1.1.1.2122.15	16.4481	37195	16.1938
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	GGICSEPKVVGPCR	Arg->Lys@14 TR135650_c2_g2_i3	cleaved A-G@N-term; missed K-V@8	0.05	1460.78	487.93	1460.74	487.92	3	18	1.1.1.2117.4	15.929	18299	15.9077
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Glu->Lys@14; Deamidated(N)@26	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	-0.02	4806.21	802.04	4806.23	802.05	6	19	1.1.1.2142.10	18.3143	53490	17.8936
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACR	Glu->Met@11	missed R-F@5; missed K-G@24	0.01	4263.92	711.66	4263.91	711.66	6	16	1.1.1.2150.4	19.0706	14087	19.1345
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.02	22.02	80.77	71.79	71.79	0	99.00	GGICSEPKVVGPCR	Lys->Arg@8	cleaved A-G@N-term; missed K-V@8	-0.01	1516.74	506.59	1516.75	506.59	3	16	1.1.1.2139.5	16.4088	29469	16.4806
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	11.57	11.57	73.08	71.79	71.79	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Lys->Glu@16	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.10	4807.23	687.76	4807.14	687.74	7	16	1.1.1.2298.8	17.6551	210995	17.7145
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Lys->Glu@16	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.07	4807.21	687.75	4807.14	687.74	7	18	1.1.1.2154.2	19.4472	8105	19.2316
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Lys->Glu@16	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.14	4807.27	687.76	4807.14	687.74	7	15	1.1.1.2174.2	21.4166	721	21.2696
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACR	Lys->Glu@24; Deamidated(N)@26	missed R-F@5; missed K-G@24	0.05	4263.89	711.66	4263.84	711.65	6	17	1.1.1.2170.3	21.0279	6329	21.0751
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Lys->Met@16	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.12	4809.26	688.04	4809.14	688.03	7	18	1.1.1.2943.2	29.2783	3362	29.1879
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Lys->Met@16	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.03	4809.17	802.54	4809.14	802.53	6	15	1.1.1.2166.5	20.6814	5595	20.5899
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Pro->Lys@17; Oxidation(M)@19	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.00	4853.23	809.88	4853.23	809.88	6	15	1.1.1.2134.15	17.5803	3120	17.5162
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACR	Pro->Thr@17	missed R-F@5; missed K-G@24	0.02	4265.93	712.00	4265.91	711.99	6	16	1.1.1.2144.3	18.4749	2782	18.5552
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	11.57	11.57	73.08	71.79	71.79	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Pro->Thr@17; Deamidated(N)@28	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.03	4811.20	688.32	4811.17	688.32	7	16	1.1.1.2299.11	17.7632	39586	17.8102
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	AAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA	Pro->Val@17	missed R-F@5; missed K-G@24; missed R-G@37; missed R-A@41	0.05	4808.26	687.90	4808.21	687.89	7	15	1.1.1.2814.2	25.9859	3139	25.992
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	10.54	10.54	88.46	88.46	73.08	0	99.00	GGICSEPKVVGPCRAAIPR	Ser->Lys@5	cleaved A-G@N-term; missed K-V@8; missed R-A@14	0.01	2038.13	510.54	2038.12	510.54	4	17	1.1.1.2142.3	16.7065	27530	16.599
TR135650_c2_g2_i1_CDS2|length	sp|B2G331|VKT1_HETCR	TRPV1	Kunitz-type	22.19	22.19	80.77	78.21	78.21	0	99.00	GGICSEPKVVGPCR	Ser->Lys@5;	cleaved A-G@N-term; missed K-V@8	0.00	1529.81	510.94	1529.81	510.94	3	14	1.1.1.2117.9	15.9498	47346	16.0051
TR145129_c1_g1_i3_CDS8|length	sp|P0DN15|VKTB_ANEVI	Kv type2	Kunitz-type	8.07	8.07	46.88	46.88	45.83	0	99.00	RCEPFIYGGCGGNANNFEDKAQCEEKCLGR	Asp->Ser@19; Lys->Arg@20	missed R-C@1; missed K-A@20; missed K-C@26	-0.01	3483.54	697.71	3483.54	697.72	5	27	1.1.1.2854.17	26.6027	7020	26.5291
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.01	1649.70	825.86	1649.69	825.85	2	17	1.1.1.2027.21	18.1955	58872	18.2009
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.01	1649.70	825.86	1649.69	825.85	2	18	1.1.1.2028.18	18.2788	58872	18.2009
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.01	1649.70	825.86	1649.69	825.85	2	18	1.1.1.2029.12	18.3794	64970	18.2009
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.70	825.86	1649.69	825.85	2	18	1.1.1.2033.11	18.7663	79726	18.2009
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	19	1.1.1.2043.5	19.7244	128832	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	19	1.1.1.2044.6	19.8206	130640	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	18	1.1.1.2045.5	19.9001	130640	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	18	1.1.1.2046.6	20.013	130640	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	17	1.1.1.2047.5	20.1247	131959	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	18	1.1.1.2048.6	20.204	128832	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.71	825.86	1649.69	825.85	2	17	1.1.1.2049.3	20.2881	131959	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.02	1649.70	825.86	1649.69	825.85	2	15	1.1.1.2053.4	20.7047	115869	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Arg->Glu@1; Tyr->Trp@3	cleaved N-S@C-term; missed R-W@1	0.01	1649.70	825.86	1649.69	825.85	2	18	1.1.1.2056.3	20.9951	54130	19.9387
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	2	99.00	RWYHGSNYGSIGSGN	Arg->Phe@1; Deamidated(N)@7	cleaved N-S@C-term; missed R-W@1	0.00	1645.67	823.84	1645.68	823.85	2	15	1.1.1.2050.5	20.3807	6045	20.4193
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Dioxidation(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	0.00	1677.70	839.86	1677.70	839.86	2	16	1.1.1.2041.7	19.5472	29194	19.5527
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Dioxidation(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	0.00	1677.70	839.86	1677.70	839.86	2	17	1.1.1.2042.7	19.6442	29194	19.5527
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Dioxidation(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	0.00	1677.70	839.86	1677.70	839.86	2	16	1.1.1.2043.6	19.741	29194	19.5527
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Dioxidation(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	0.00	1677.70	839.86	1677.70	839.86	2	15	1.1.1.2045.6	19.9167	29194	19.5527
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	His->Met@4	cleaved N-S@C-term; missed R-W@1	-0.02	1647.69	824.85	1647.71	824.86	2	18	1.1.1.2036.7	19.0654	64339	19.0714
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	His->Met@4	cleaved N-S@C-term; missed R-W@1	-0.02	1647.69	824.85	1647.71	824.86	2	17	1.1.1.2037.8	19.1615	64339	19.0714
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	His->Met@4	cleaved N-S@C-term; missed R-W@1	-0.02	1647.69	824.85	1647.71	824.86	2	18	1.1.1.2038.7	19.2448	64339	19.0714
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	His->Met@4	cleaved N-S@C-term; missed R-W@1	-0.02	1647.69	824.85	1647.71	824.86	2	17	1.1.1.2039.6	19.3537	64339	19.0714
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	His->Met@4; Deamidated(N)@7	cleaved N-S@C-term; missed R-W@1	0.00	1648.69	825.35	1648.69	825.35	2	13	1.1.1.2040.5	19.4505	81836	19.0714
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Trp->Kynurenin(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	0.00	1649.71	825.86	1649.71	825.86	2	19	1.1.1.2042.6	19.6275	130640	19.6496
TR148434_c0_g2_i1_CDS3|length	sp|Q9U6X1|ACTP3_HETMG	Cytolysin type2	Cytolysin	6	6	30.16	30.16	30.16	0	99.00	RWYHGSNYGSIGSGN	Trp->Kynurenin(W)@2; His->Glu@4	cleaved N-S@C-term; missed R-W@1	-0.01	1649.70	825.86	1649.71	825.86	2	16	1.1.1.2057.4	21.0915	29561	20.035
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Asp->Asn@9; Dioxidation(W)@17		-0.02	2041.91	681.64	2041.93	681.65	3	15	1.1.1.2851.13	26.2991	8714	26.2421
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Gly->Thr@6		-0.02	2054.93	685.98	2054.95	685.99	3	19	1.1.1.2858.14	26.9733	12621	26.4334
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	2	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	19	1.1.1.2884.2	29.4074	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	18	1.1.1.2880.3	29.0449	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	19	1.1.1.2882.2	29.22	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	19	1.1.1.2883.2	29.3137	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	19	1.1.1.2885.2	29.5011	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	18	1.1.1.2886.2	29.5948	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	18	1.1.1.2888.2	29.7823	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	17	1.1.1.2881.2	29.1262	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	16	1.1.1.2887.2	29.6887	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	CASFYTIFADCCRRPR	Phe->Leu@8	missed R-R@13	0.01	2005.96	502.50	2005.95	502.50	4	14	1.1.1.2879.3	28.9412	107714	29.2106
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Thr->Asn@7; Val->Met@8		0.00	2055.89	686.30	2055.89	686.30	3	18	1.1.1.2851.15	26.3074	6401	26.2421
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Thr->Asp@7		0.00	2024.90	675.97	2024.90	675.97	3	16	1.1.1.2854.15	26.5944	9874	26.2421
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Val->Asn@8		0.02	2025.91	676.31	2025.89	676.31	3	17	1.1.1.2851.12	26.2949	11274	26.2421
TR18604_c0_g1_i1_CDS1|length; TR130412_c0_g1_i1_CDS3|length	sp|Q76CA0|NA2G3_STIGI	Nav type2	Defensin	19.34	19.34	41.67	41.67	41.67	0	99.00	SATLTGTVDLGSCNEGWEK	Val->Pro@8; Oxidation(W)@17		-0.01	2024.89	1013.45	2024.90	1013.46	2	16	1.1.1.2850.21	26.2369	278	26.1465
TR69080_c0_g1_i1_CDS11|length	sp|B1B5J0|SHTX5_STIHA	EGF-like sea anemone	EGF-like	2.5	2.5	26.74	26.74	26.74	0.5	99.00	RDEGVACTGQHASSFCLNGGTCR	Arg->Phe@1	missed R-D@1	0.01	2491.07	831.36	2491.06	831.36	3	21	1.1.1.2303.19	16.7408	61081	16.7544
TR69080_c0_g1_i1_CDS11|length	sp|B1B5J0|SHTX5_STIHA	EGF-like sea anemone	EGF-like	8	8	56.98	55.81	55.81	0	99.00	RDEGVACTGQHASSFCLNGGTCR	Arg->Phe@1	missed R-D@1	-0.03	2491.03	831.35	2491.06	831.36	3	15	1.1.1.2164.11	20.4875	4580	20.396
TR69080_c0_g1_i1_CDS11|length	sp|B1B5J0|SHTX5_STIHA	EGF-like sea anemone	EGF-like	8	8	56.98	55.81	55.81	0	99.00	HIASLGEYYCICPGDYTGHRCDQKS	Asp->Asn@22	cleaved S-G@C-term; missed R-C@20; missed K-S@24	-0.07	2946.24	983.09	2946.31	983.11	3	19	1.1.1.2125.19	16.7464	46134	16.76
TR69080_c0_g1_i1_CDS11|length	sp|B1B5J0|SHTX5_STIHA	EGF-like sea anemone	EGF-like	8	8	56.98	55.81	55.81	0	99.00	HIASLGEYYCICPGDYTGHRCDQKS	Asp->Asn@22	cleaved S-G@C-term; missed R-C@20; missed K-S@24	-0.07	2946.24	983.09	2946.31	983.11	3	20	1.1.1.2126.20	16.8444	46134	16.76
TR71631_c0_g6_i2_CDS2|length; TR71631_c0_g6_i1_CDS1|length	sp|B1P1G9|JZ20A_CHIGU	ICK - Unknow	ICK	4	4	44.05	23.81	23.81	2	99.00	GLACYPDPYESGRYVCSKER	Glu->Gln@19	missed R-Y@13	-0.01	2379.09	595.78	2379.10	595.78	4	16	1.1.1.1536.4	16.899	30613	16.8783
TR71631_c0_g6_i2_CDS2|length; TR71631_c0_g6_i1_CDS1|length	sp|B1P1G9|JZ20A_CHIGU	ICK - Unknow	ICK	4	4	44.05	23.81	23.81	0	99.00	GLACYPDPYESGRYVCSKER	Glu->Gln@19	missed R-Y@13	-0.01	2379.09	595.78	2379.10	595.78	4	15	1.1.1.1534.6	16.7149	30613	16.8783
TR71631_c0_g6_i2_CDS2|length; TR71631_c0_g6_i1_CDS1|length	sp|B1P1G9|JZ20A_CHIGU	ICK - Unknow	ICK	4	4	44.05	23.81	23.81	0	99.00	GLACYPDPYESGRYVCSKER	Glu->Gln@19	missed R-Y@13	-0.01	2379.09	595.78	2379.10	595.78	4	16	1.1.1.1535.6	16.8106	30613	16.8783
TR71631_c0_g6_i2_CDS2|length; TR71631_c0_g6_i1_CDS1|length	sp|B1P1G9|JZ20A_CHIGU	ICK - Unknow	ICK	4	4	44.05	23.81	23.81	0	99.00	GLACYPDPYESGRYVCSKER	Glu->Gln@19	missed R-Y@13	-0.01	2379.09	595.78	2379.10	595.78	4	17	1.1.1.1537.3	16.9917	30613	16.8783
TR72284_c0_g1_i1_CDS4|length	sp|A7RMN1|KV51_NEMVE	Kv Type5	Kv Type5	13.14	13.14	66.67	62.67	62.67	2	99.00	KYQRCTSNSQCCKTPKDFAGRTLR	Ser->Gly@7	missed K-Y@1; missed R-C@4; missed K-T@13; missed R-T@21	0.05	2892.47	579.50	2892.42	579.49	5	20	1.1.1.2113.5	15.5574	50996	15.524
TR72284_c0_g1_i1_CDS4|length	sp|A7RMN1|KV51_NEMVE	Kv Type5	Kv Type6	13.14	13.14	66.67	62.67	62.67	0	99.00	KYQRCTSNSQCCKTPKDFAGRTLR	Ser->Gly@7	missed K-Y@1; missed R-C@4; missed K-T@13; missed R-T@21	0.05	2892.47	579.50	2892.42	579.49	5	18	1.1.1.2112.8	15.4687	50996	15.524
TR72284_c0_g1_i1_CDS4|length	sp|A7RMN1|KV51_NEMVE	Kv Type5	Kv Type7	13.14	13.14	66.67	62.67	62.67	0	99.00	KYQRCTSNSQCCKTPKDFAGRTLR	Ser->Gly@7	missed K-Y@1; missed R-C@4; missed K-T@13; missed R-T@21	0.12	2892.53	483.10	2892.42	483.08	6	16	1.1.1.2113.2	15.5365	51163	15.524
TR96729_c4_g2_i1_CDS1|length; TR96729_c4_g1_i3_CDS26|length	sp|Q9TWG1|K1B_ANESU	Kv type1	ShKT domain	6.01	6.02	32.43	32.43	32.43	0	99.00	ACKDSLPSHICQNVKQNHSCGSEKYKTHCAKTCGAC	Gln->Met@16	missed K-Q@15; missed K-Y@24; missed K-T@26; missed K-T@31	0.05	4145.91	691.99	4145.86	691.98	6	15	1.1.1.1518.17	15.2127	109928	14.8523
TR96729_c4_g2_i1_CDS1|length; TR96729_c4_g1_i3_CDS26|length	sp|Q9TWG1|K1B_ANESU	Kv type1	ShKT domain	6.01	6.02	32.43	32.43	32.43	0	99.00	ACKDSLPSHICQNVKQNHSCGSEKYKTHCAKTCGAC	Gln->Met@16	missed K-Q@15; missed K-Y@24; missed K-T@26; missed K-T@31	0.05	4145.91	691.99	4145.86	691.98	6	15	1.1.1.1518.17	15.2127	109928	14.8523

β-defensin - Na$_V$ Type 2

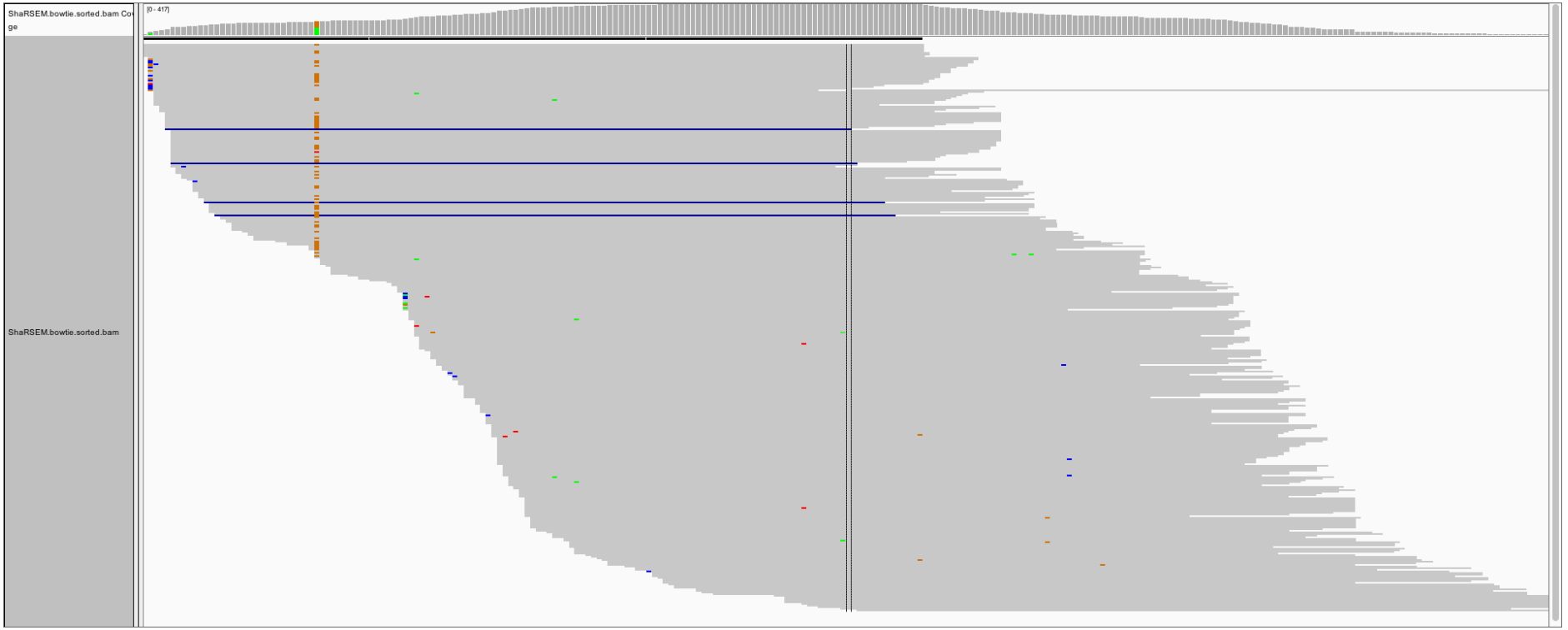
TR18604_c0_g1_i1



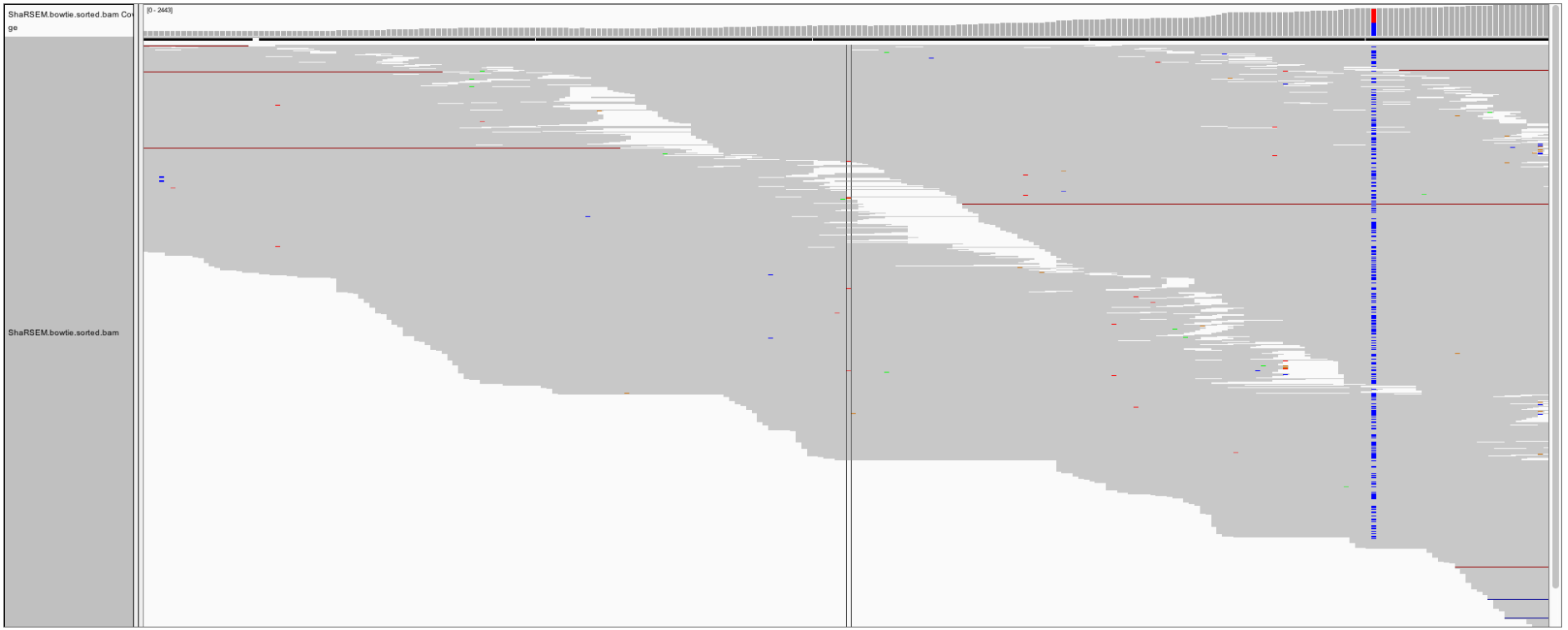

20
40
60
80
TR18604_c0_g1_i1 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTIFADCCRRPRG 84
TR130412_c0_g1_i1 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTIFADCCRRPRG 84
TR18604_c0_g1_i2 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTILADCCRRPRG 84
TR130412_c0_g1_i2 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTILADCCRRPRG 84
100%
Conservation
0%

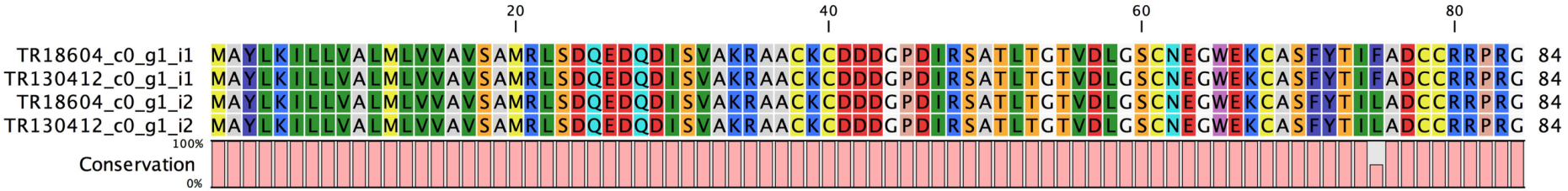
β-defensin - Nav Type 2

TR18604_c0_g1_i1



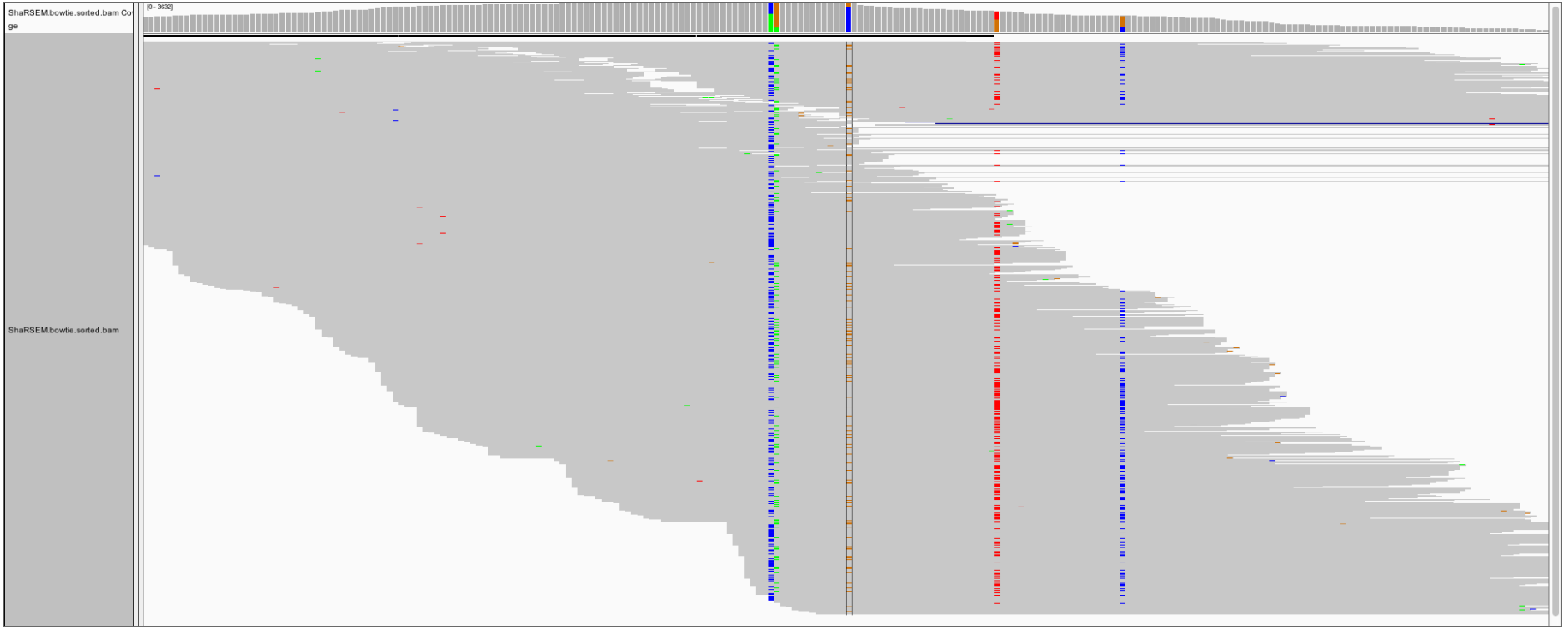
TR130412_c0_g1_i1

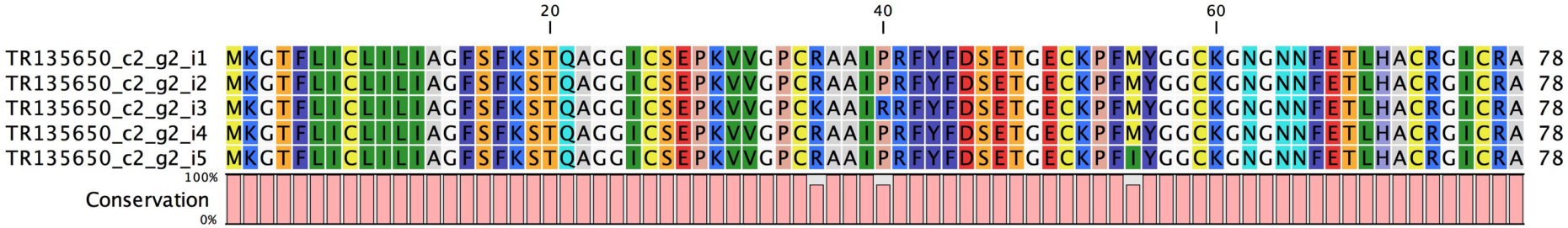



20
40
60
80
TR18604_c0_g1_i1 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTIFADCCRRPRG 84
TR130412_c0_g1_i1 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTIFADCCRRPRG 84
TR18604_c0_g1_i2 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTILADCCRRPRG 84
TR130412_c0_g1_i2 MAYLKILLVALMLVVAVSAMRLSDQEDQDISVAKRAACKCDDDGPDIRSATLTGTVDLGSCNEGWEKCASFYTILADCCRRPRG 84
100%
Conservation
0%

Kunitz-TRPV1

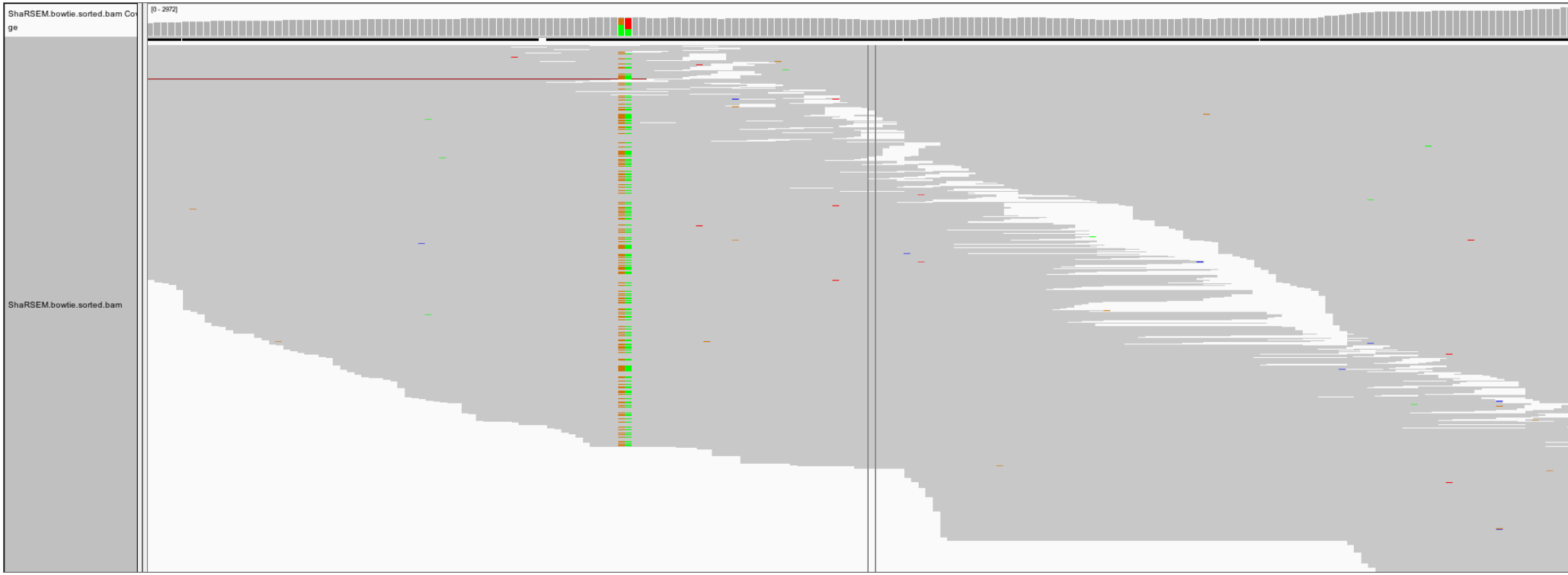
TR135650_c2_g2_i1




20
40
60
TR135650_c2_g2_i1 MKGTFLICLILIAGFSFKSTQAGGICSEPKVVGPCRAAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA 78
TR135650_c2_g2_i2 MKGTFLICLILIAGFSFKSTQAGGICSEPKVVGPCRAAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA 78
TR135650_c2_g2_i3 MKGTFLICLILIAGFSFKSTQAGGICSEPKVVGPCKAAIRRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA 78
TR135650_c2_g2_i4 MKGTFLICLILIAGFSFKSTQAGGICSEPKVVGPCRAAIPRFYFDSETGECKPFMYGGCKGNGNNFETLHACRGICRA 78
TR135650_c2_g2_i5 MKGTFLICLILIAGFSFKSTQAGGICSEPKVVGPCRAAIPRFYFDSETGECKPFIYGGCKGNGNNFETLHACRGICRA 78
100%
Conservation
0%

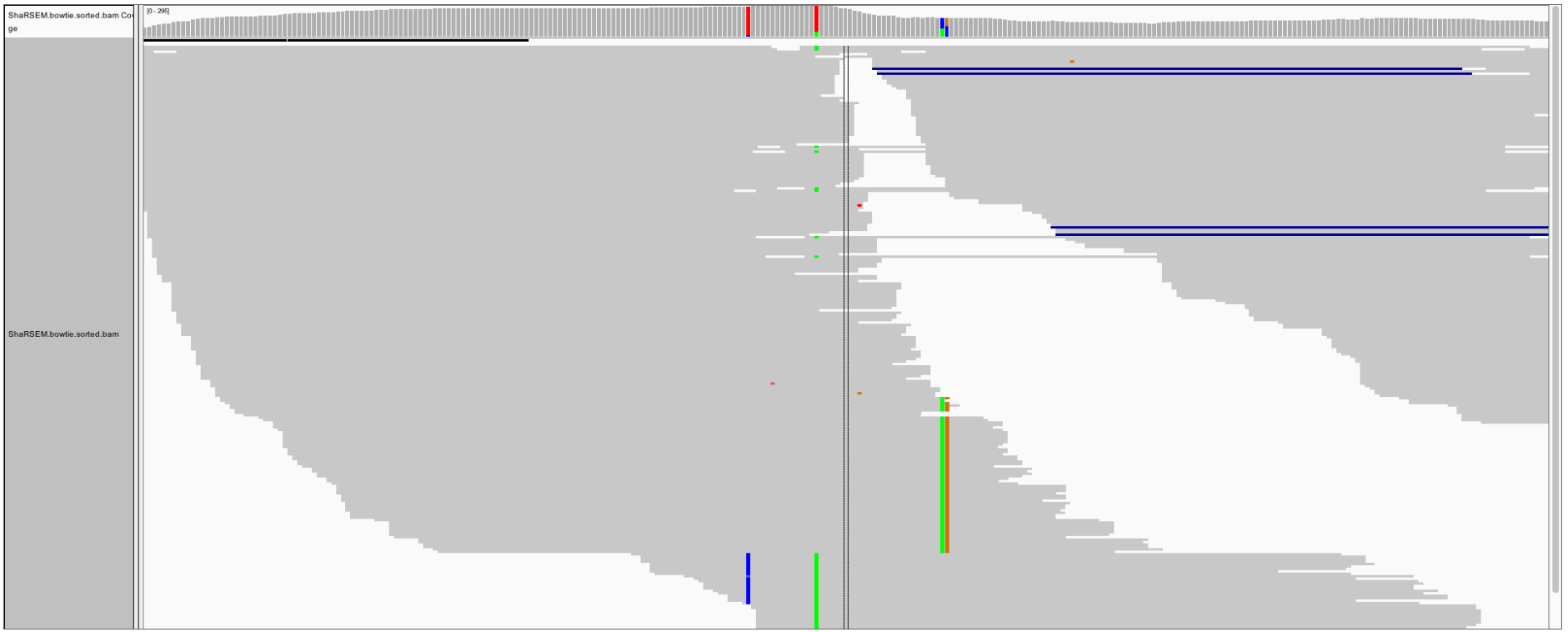
K_V type 4 / sea anemone structural class 9a

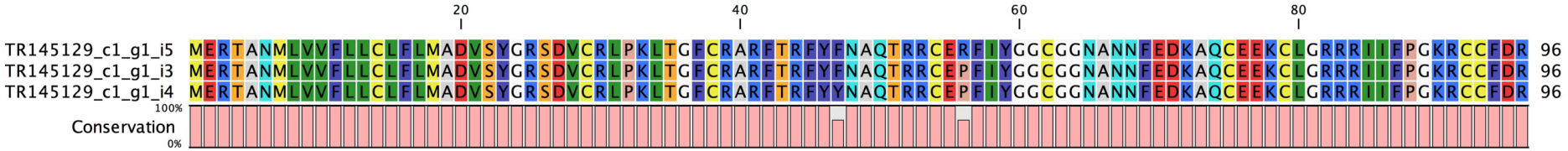
TR138219_g4_c1_i1



Kunitz – Kv type 2

TR145129_c1_g1_i3

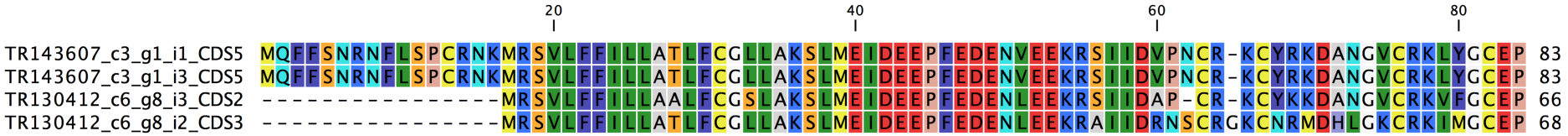



20
40
60
80
TR145129_c1_g1_i5 MERTANMLVVFLLCLFLMADVSYGRSDVCRLPKLTGFCRARFTRFYFNAQTRRCERFIYGGCGGNANNFEDKAQCEEKCLGRRRIIFPGKRCCFDR 96
TR145129_c1_g1_i3 MERTANMLVVFLLCLFLMADVSYGRSDVCRLPKLTGFCRARFTRFYFNAQTRRCEPFIYGGCGGNANNFEDKAQCEEKCLGRRRIIFPGKRCCFDR 96
TR145129_c1_g1_i4 MERTANMLVVFLLCLFLMADVSYGRSDVCRLPKLTGFCRARFTRFYYNAQTRRCEPFIYGGCGGNANNFEDKAQCEEKCLGRRRIIFPGKRCCFDR 96
100%
Conservation
0%

Appendix IV .Sequence alignment of putative toxins with no matches to any database entries.

Alignment of representative sequences from the putative toxin families U_1-Std to U_{12}-Std

U_1-Std



U_2-Std

U_3-Std

U_4-Std

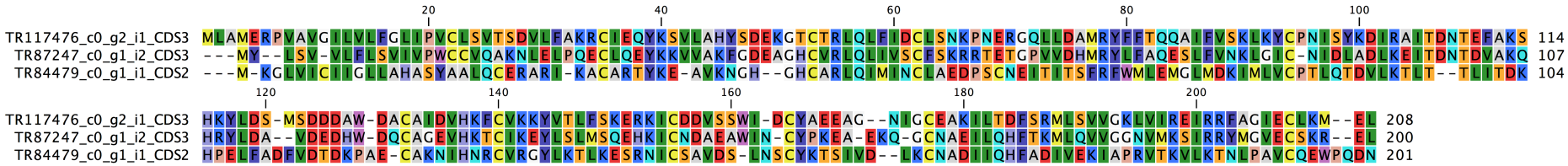
U_5-Std

U_6-Std

U_7-Std

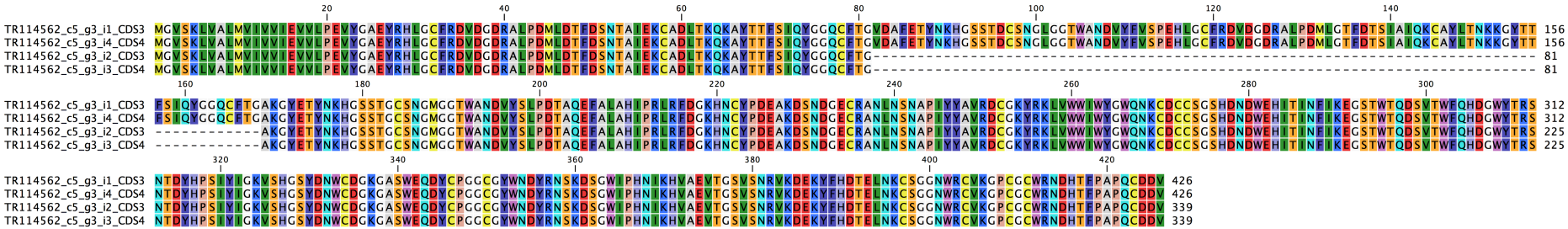
U_8-Std

U_9-Std



U_{10}-Std

U_{11}-Std



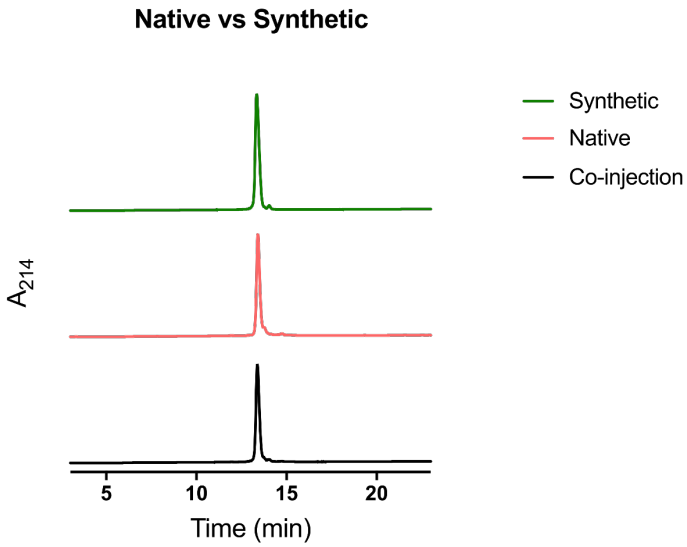
U_{12}-Std

Appendices of Chapter 3

Appendix V Hidden polymorphism within the Trinity assembled Ate1a-encoding contig.

Appendix Figure V.i – Compressed global view of the CDS region of TR5054|c0_g1_i1 showing mapped quality trimmed reads with bases identical to the contig in grey and polymorphisms in colour. Red square in top panel indicates viewed portion of the full length contig, while vertical coloured bars in the read coverage display indicates relative frequencies of polymorphisms.

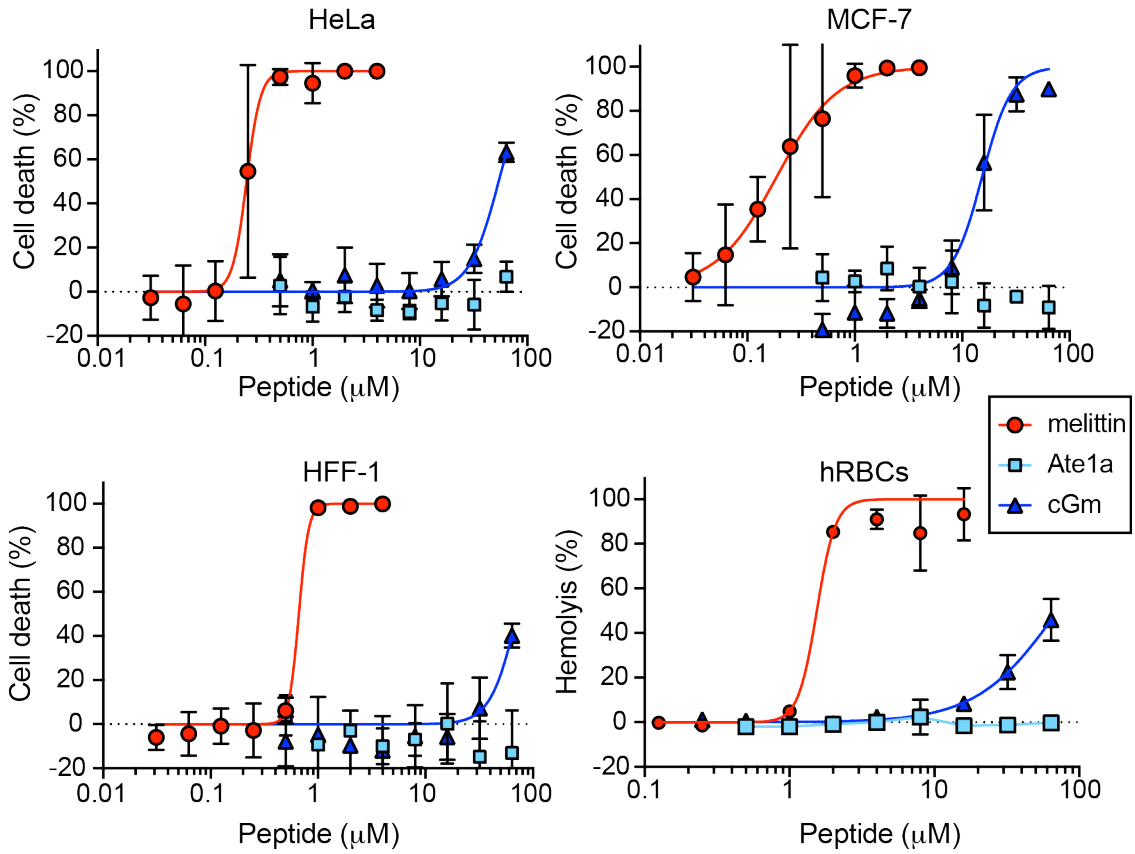
Appendix VI Co-elution of native and synthetic Ate1a



Appendix Figure VI.i – RP-HPLC chromatograms showing the retention times for native Ate1a, synthetic Ate1a, and a mixture of native and synthetic Ate1a (co-injection).

Appendix VII Ate1a is neither cytotoxic nor haemolytic

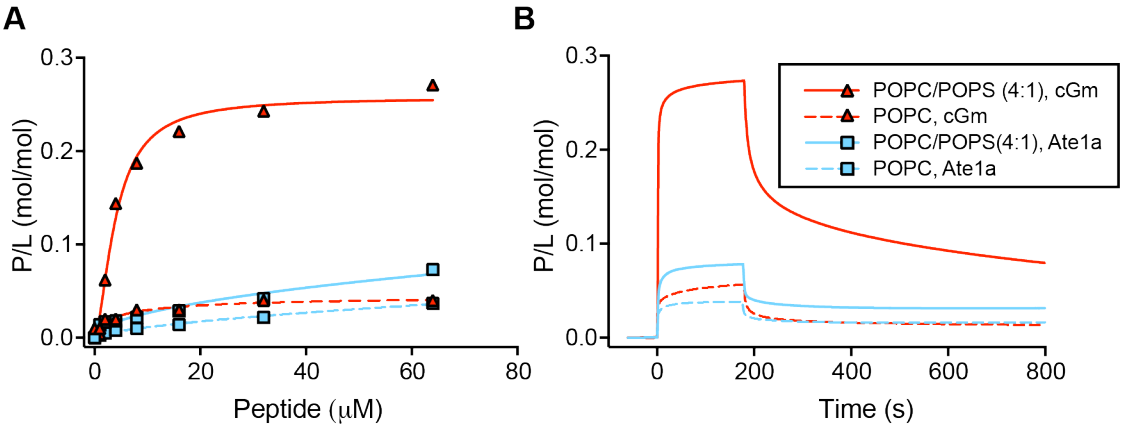
Ate1a was not toxic against tested cultured cell lines (HeLa, human cervical adenocarcinoma cells; MCF-7, human breast adenocarcinoma; and HFF-1, human foreskin fibroblast non-cancerous cells), and did not causes hemolysis of human red blood cells (hRBCs) up to 64 μM. Melittin, a bee venom-peptide known to disrupt cell membranes, and a cyclic version of gomesin (cGm), an antimicrobial peptide from spider hemocytes, were included as controls. Melittin shows toxicity at low micromolar concentrations, whereas cGm shows mild toxicity against all tested cells.



Appendix Figure VII.i – Toxicity induced by Ate1a against HeLa, MCF-7, HFF-1 as measured by a resazurin assay, and hemolysis induced against human red blood cells, as measured by release of haemoglobin. Melittin and cyclic Gomesin (cGm) were included as controls. Data points are mean ± SD of three independent replicates.

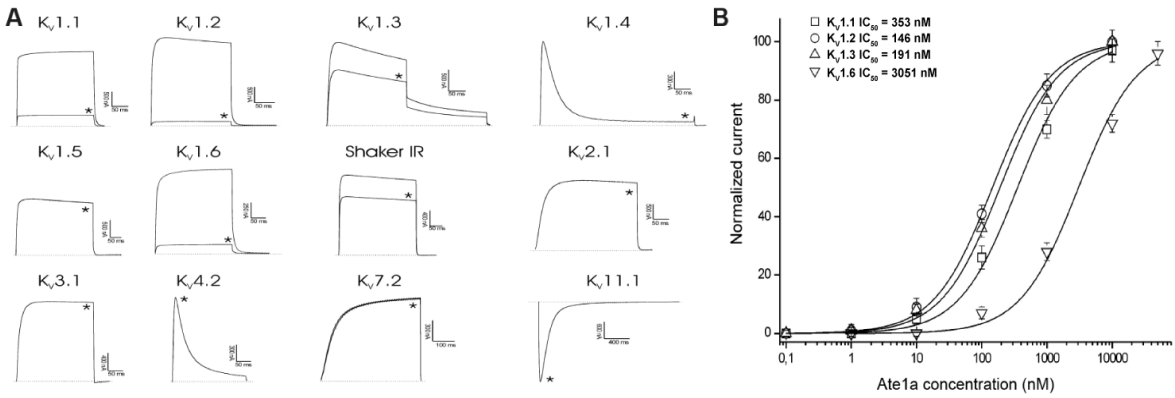
Appendix VIII Ate1a interacts only weakly with lipid bilayers

The ability of Ate1a to bind to lipid bilayers was examined by Surface Plasmon Resonance (SPR). Ate1a has weak affinity for neutral (POPC) membranes, only slightly improved with negatively-charged POPC/POPS (4:1) membranes, as shown by the dose response curves (Appendix Figure VIII.iA), and by sensorgrams obtained with 64 μM (Appendix Figure VIII.i). cGm, included for comparison, shows high affinity for negatively-charged model membranes.



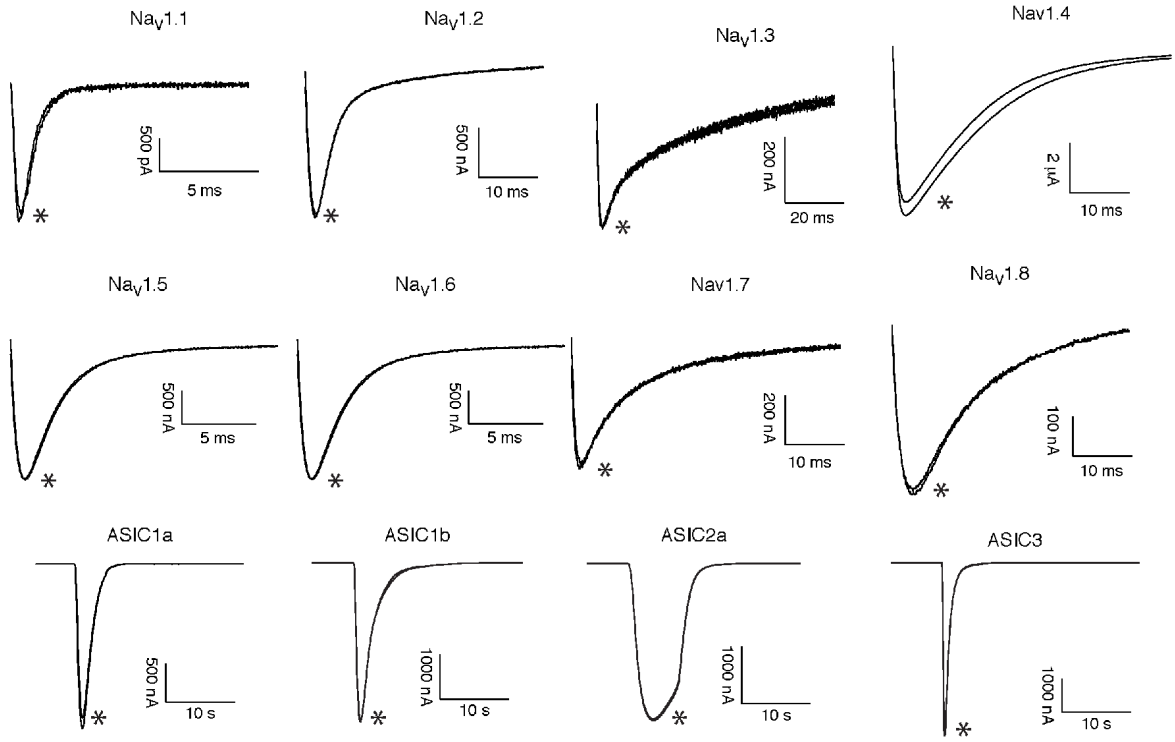
Appendix Figure VIII.i – Binding of Ate1a to model membranes as examined by surface plasmon resonance. Peptide samples were injected for 180 s (association phase) over lipid bilayers composed of POPC or POPC/POPS (4:1 molar ratio) deposited onto an L1 chip. Dissociation from the membrane was followed for 600 s (dissociation phase). Response units (RU) were converted into peptide-to-lipid ratio (P/L (mol/mol) by converting RUs into amount of peptide and normalized to the amount of lipid deposited onto the chip surface (1 RU = 1pg/mm^2 of lipid or peptide). cGm was included as control. (A) Dose-response curves using a reporting point at the end of association phase (injection at t = 170 s). **(B)** Sensorgrams obtained upon injection of peptide at 64 μM.

Appendix IX Electrophysiological characterization of K_V isoforms inhibited by Ate1a



Appendix Figure IX.i – Electrophysiological characterization of K_V isoforms inhibited by Ate1a. (A) Representative whole-cell current traces obtained from K_V channels expressed in *Xenopus* oocytes in the absence (control) and presence (*) of 3 µM Ate1a. **(B)** Concentration-response curves obtained by plotting current inhibition as a function of increasing Ate1a concentration.

Appendix X Ate1a is not active on Na_V and ASIC channels



Appendix Figure X.i – Representative whole-cell current traces obtained from Na_V or ASIC channels expressed in *Xenopus* oocytes in the absence (control) and presence (*) of 3 µM Ate1a. Ate1a has no significant effect on any of these channels.

Appendix XI Structural statistics for the ensemble of Ate1a

Appendix Table XI.i Structural statistics for the ensemble of Ate1a[a]

Experimental restraints	
Inter-proton distance restraints	
Total	66
Intra-residue ($i = j$)	21
Sequential ($\|i - j\| = 1$)	37
Medium range ($1 < \|i - j\| < 5$)	8
Long range ($\|i - j\| \geq 5$)	0
Disulfide bond restraints	6
Dihedral-angle restraints	
ϕ dihedral angle restraints	11
ψ dihedral angle restraints	13
χ^1 angle restraints	1
Total number of restraints per residue	5.4
Violations of experimental restraints	0
RMSD from mean coordinate structure (Å)[b]	
All backbone atoms	1.11 ± 0.26
All heavy atoms	2.06 ± 0.39
Stereochemical quality[c]	
Ramachandran plot statistics	
Residues in most favored Ramachandran region (%)	95.0 ± 4.8
Disallowed regions (%)	0.0 ± 0.0
Unfavorable sidechain rotamers (%)	7.5 ± 6.7
Clashscore, all atoms[d]	0.0 ± 0.0
Overall MolProbity score	1.34 ± 0.38

[a] All statistics are given as mean ± S.D.

[b] Mean RMSD calculated over the entire ensemble of 20 structures.

[c] Sterochemical quality according to MolProbity (http://helix.research.duhs.duke.edu).

[d] Clashscore is defined the number of steric overlaps >0.4 Å per 1000 atoms.

Appendix XII Antimicrobial assay: SI_Ate1a_Antimicrobial.xlsx

Appendix Table XII.i Antimicrobial activity of Ate1a toward five bacteria: *Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*, and two yeast: *Candida albicans* and *Cryptococcus neoformans*.

CompoundID	CompoundName	ProjectID	RunID	Hit	Tox	Sa (MSSA)	Sa	Ec	Kp	Pa	Ab	Ca	Cn	Hk	Unit	Media	Plate size
C0303419	Ate1a	P0426	HVR00067	0	0		>32	>32	>32	>32	>32	>32	>32	>32	ug/mL	CA-MHB	384-w
Cooper data	Gomesin					32		4	32	8	4-8	16	1			MHB	384-w
Cooper data	Gomesin						128	16	64	16	16	16	2	>100		MHB	96-w