



Thèse

2015

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Venomics of the piscivorous cone snail *C. consors*: multidisciplinary techniques for drug discovery

Biasse, Daniel

How to cite

BIASSE, Daniel. Venomics of the piscivorous cone snail *C. consors*: multidisciplinary techniques for drug discovery. Doctoral Thesis, 2015. doi: 10.13097/archive-ouverte/unige:75353

This publication URL: <https://archive-ouverte.unige.ch/unige:75353>

Publication DOI: [10.13097/archive-ouverte/unige:75353](https://doi.org/10.13097/archive-ouverte/unige:75353)

UNIVERSITE DE GENEVE
Sciences pharmaceutiques

FACULTE DES SCIENCES
Prof. Denis Hochstrasser

ATHERIS LABORATORIES

Dr. Reto Stöcklin

**Venomics of the piscivorous cone snail *C. consors*:
multidisciplinary techniques for drug discovery**

THESE

présentée à la faculté des sciences de l'Université de Genève

pour obtenir le grade de Docteur ès sciences,

mention interdisciplinaire

par

Daniel Biasse

de

France

Thèse N°4808

Genève

Centre d'impression UNIGE

2015



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

***Doctorat ès sciences
Mention interdisciplinaire***

Thèse de ***Monsieur Daniel Herbert BIASSE***

intitulée :

**"Venomics of the Piscivorous Cone Snail *C. consors*:
Multidisciplinary Techniques for Drug Discovery"**

La Faculté des sciences, sur le préavis de Monsieur D. HOCHSTRASSER, professeur ordinaire et directeur de thèse (Faculté de médecine, Département de médecine génétique et de laboratoire et Section des sciences pharmaceutiques), Monsieur R. STÖCKLIN, docteur et codirecteur de thèse (Atheris SA, Arare, Suisse), Monsieur J.-L. WOLFENDER, professeur ordinaire (Section des sciences pharmaceutiques), Madame F. LISACEK, docteure (Institut Suisse de Bioinformatique), Monsieur E. ALLÉMANN, professeur ordinaire (Section des sciences pharmaceutiques) et Monsieur J. J. CALVETE, professeur (Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas, Valencia, España), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 10 juillet 2015

Thèse - 4808 -

Le Décanat

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

To my family

Abstract

Venomous animals have always been intriguing, whether they scare or captivate people. For scientists, they have always been fascinating creatures not only for their biological aspects, but also for the power of their venom. Venomous creatures can be found in virtually all major groups of terrestrial and marine life with more than a 100,000 species spread throughout the world. Scientists have always wondered what can be so powerful in such a tiny drop of venom that can cause such damage and possibly kill. Only quite recently, in comparison to the millions of years of evolution of venomous animals, have scientists started in-depth investigations on how such destructive power could be used in a constructive manner. Progressively, the number of venom components studied increased and so did their potential as drug candidates for healthcare through pharmacological research and development.

In the early 2000s, venoms were studied in a more and more comprehensive manner. Venom proteomics was launched. With the advent of next-generation sequencing in molecular biology, the use of this technology in the study of venom glands led to venom transcriptomics. Then, studies combining these different technologies were developed towards integrated venom studies: venomomics was born. The *raison d'être* of venomomics is not only to better understand the venomous function purely from a biological slant, but also to understand the evolution and diversity of venom peptides and proteins by using a combination of genomics, transcriptomics and proteomics. Of course, one of the major goals of venomomics projects is also the development of new compounds for pharmacological research and drug discovery. These developments have emerged at the right time, just when the pharmaceutical industry – following the failure of combinatorial chemistry to hold its promises – is switching back to natural or natural-derived compounds for their future applications in healthcare.

It is in this context that the CONCO project – the cone snail genome project for health – was launched using a piscivorous cone snail, *Conus consors*, as a model. The work reported in this thesis covers part of the proteomics, peptidomics and transcriptomics studies undertaken during the CONCO project.

After an introduction on some biological aspects of cone snails including their venomous apparatus and of course their venom, the first part of this thesis describes the extensive study of the peptide content of *C. consors* venom – i.e. conopeptides – by using different mass spectrometry techniques and by optimising the sampling and analytical methods to obtain the largest possible dataset. The number of components in cone snail venom soon appeared largely underestimated, pointing to the necessity of using several analytical techniques to cover the high diversity of conopeptides in terms of structure and physico-chemical properties. The importance of sampling methods is also herewith demonstrated and contributes to the complexity of cone snail venom studies.

The second part of this work focuses on the study of the intermediate source of conopeptide: mRNA transcripts. Until recently, little was known of conopeptide processing. Many important steps occur in the transformation of gene sequences into final peptide products. Classical RNA probing virtually entailed knowing what to look for. The methods were time consuming and not well suited to the analysis of the whole transcript content of a given tissue. With the advent of next-generation sequencing, its application in venomomics did not take long. Using the latest sequencing technology, the transcriptome of *C. consors* brought for the first time a comprehensive view of the venom duct transcripts, giving access not only to known families of conopeptides, but also to rare and hitherto unknown conopeptide and conoprotein transcripts and other elements “recruited” for the venomous function.

Finally, the last part covers the characterisation of a maximum of *C. consors* venom peptides and proteins. For this purpose, different sampling and analytical methods were used, but the

most important factor was the introduction of transcriptomics data in the workflow - thanks to bioinformatics - to assist in the peptidomics and proteomics content deconvolution and characterisation. Even though the complex post-translational modifications of conopeptides or conoproteins still restrain the sequence data output, many new conopeptide or conoprotein families were characterised. Thanks to this integrative approach, the different sequence datasets put side by side provide a more complete picture of the venom production and maturation processes that take place in the venom gland of a cone snail. This approach is now a standard in venomomics and could in the future be integrated in the workflow of new biological assays, like microfluidic chip-based assays, to speed up the research and development of new pharmacological tools and drug candidates for healthcare.

Résumé

Les animaux venimeux ont toujours été intrigants, qu'ils effrayent ou qu'ils attirent les gens. Pour les scientifiques, ils ont toujours été fascinants, non seulement de par leurs caractéristiques biologiques, mais aussi à cause de la puissance de leur venin. Les animaux venimeux peuvent être trouvés dans pratiquement tous les grands groupes terrestres ou marins avec plus de 100'000 espèces réparties dans le monde. Les scientifiques se sont toujours demandé ce qui pouvait être si puissant dans une petite goutte de venin et causer autant de dégâts, ou même la mort. Ce n'est que récemment, à l'échelle des millions d'années d'évolution des animaux venimeux, que les scientifiques ont entrepris des recherches plus poussées pour déterminer comment une puissance aussi destructive pourrait être utilisée de manière constructive. Le nombre de composés de venin étudiés a progressivement augmenté, ainsi que leur potentiel en tant que médicament pour la santé grâce à la recherche et au développement pharmacologique.

Au début des années 2000, les venins ont été de plus en plus étudiés, et ce de manière plus globale. La protéomique des venins fut lancée. Avec l'arrivée en biologie moléculaire du séquençage de deuxième génération, l'utilisation de cette technologie dans l'étude des glandes à venin a naturellement mené à la transcriptomique des venins. Puis, les études combinant ces différentes technologies ont évolué vers des études intégrées sur les venins : ce fut la naissance de la vénomique. La raison d'être de la vénomique n'est pas seulement de mieux comprendre la fonction venimeuse d'un point de vue purement biologique, mais aussi de comprendre l'évolution et la diversité des peptides et protéines de venin grâce à la génomique, la transcriptomique et la protéomique. Evidemment, l'un des buts majeurs des projets en vénomique est aussi le développement de nouveaux composés pour la recherche pharmacologique et la découverte de médicaments. Ces développements sont arrivés à

point nommé, au moment où l'industrie pharmaceutique – après l'échec de la chimie combinatoire à tenir ses promesses – est en train de revenir à l'utilisation de composés naturels ou de dérivés naturels pour de futures applications dans le domaine de la santé.

C'est dans ce contexte que le projet CONCO – « The cone snail genome for health » – fut lancé en utilisant un cône piscivore, *Conus consors*, comme modèle. Le travail rapporté dans cette thèse couvre une partie des études sur la protéomique, peptidomique et transcriptomique entreprises pendant le projet CONCO.

Après une introduction sur quelques aspects biologiques des cônes, notamment leur appareil venimeux et leur venin bien sûr, la première partie de cette thèse décrit l'étude exhaustive du contenu peptidique du venin de *C. consors* – c.-à-d. les conopeptides – grâce à différentes techniques de spectrométrie de masse et à l'optimisation de l'échantillonnage et des méthodes analytiques afin d'obtenir une plage de données la plus large possible. Le nombre de composés dans le venin de cône est rapidement apparu comme largement sous-estimé, dictant la nécessité d'utiliser plusieurs techniques analytiques pour couvrir la diversité des conopeptides en termes de structure et de propriétés physico-chimiques. L'importance des méthodes d'échantillonnage est ici aussi démontrée et contribue à la complexité des études sur le venin de cône.

La deuxième partie de ce travail met l'accent sur l'étude de la source intermédiaire des conopeptides : les transcrits ARNm. Jusqu'à récemment, le procédé de fabrication des conopeptides n'était que peu connu. Plusieurs étapes importantes ont lieu dans la transformation des séquences génomiques en composés peptidiques finaux. La méthode classique de sonde ARN impliquait quasiment de savoir ce que l'on cherchait. Ces méthodes étaient longues et peu adaptées à l'analyse d'un contenu complet des transcrits d'un tissu donné. Avec l'arrivée du séquençage de deuxième génération, son application dans la vénomique ne fut pas longue. Avec l'utilisation de la dernière technologie en séquençage, le

transcriptome du *C. consors* proposa pour la première fois une vue plus complète des transcrits présents dans le conduit à venin, donnant accès non seulement à des familles de conopeptides connues, mais également à des transcrits de conopeptides et conoprotéines rares et jusqu'ici inconnus, et également à d'autres éléments « recrutés » pour la fonction venimeuse.

Enfin, la dernière partie couvre la caractérisation d'un maximum de peptides et protéines du venin du *C. consors*. A cet effet, différentes méthodes d'échantillonnage et analytiques furent utilisées, mais le facteur décisif fut l'introduction des données transcriptomiques dans le processus – grâce à la bioinformatique – pour assister la deconvolution et la caractérisation du contenu peptidique et protéomique. Même si les modifications post-traductionnelles complexes des conopeptides et conoprotéines réfèrent toujours l'obtention de données de séquence, plusieurs nouvelles familles de conopeptides ou conoprotéines furent caractérisées. Grâce à cette approche intégrée, les différents jeux de données mis côte à côte fournissent une vue plus complète des processus de fabrication et maturation du venin qui ont lieu dans la glande à venin d'un cône. Cette approche est maintenant devenue un standard dans la vénomique et pourrait être intégrée dans les procédés de nouveaux essais biologiques, comme les essais micro-fluidiques basées sur micropuce, pour accélérer la recherche et développement de nouveaux outils pharmacologiques et de candidats-médicaments pour la santé.

Table of Contents

Abstract.....	1
<i>Résumé</i>	5
Table of Contents.....	9
List of Figures.....	13
List of Tables.....	17
Abbreviations	19
Publications.....	21
Articles	21
Presentation and posters.....	22
1. Venomics	25
1.1 Introduction.....	25
1.2 Venomous animals and their venoms	25
1.3 From natural products to drug candidates.....	26
1.4 What about venom compounds as drug candidates?.....	28
1.5 The CONCO project	29
1.6 Thesis work	30
1.7 Reference list.....	33
2. Cone snails	35
2.1 Introduction.....	35
2.2 Biology of the cone snail.....	35
2.2.1 General morphology	35
2.2.2 Venom apparatus	38
2.2.3 Habitats and feeding habits.....	43
2.3 Reproduction and Development of the cone snail.....	44

2.4 Evolution of the classification	47
2.5 Reference list.....	48
3. Conopeptides and Conoproteins	51
3.1 Introduction.....	51
3.2 Conopeptides biosynthesis and classification	51
3.3 Conopeptides nomenclature	54
3.4 Conopeptides PTMs	55
3.5 Conopeptides targets.....	57
3.6 Reference List	59
4. Cone snail venom content studies	63
4.1 Introduction: From the first case of poisoning to the study of conotoxins.....	63
4.2 The early stages of peptidomics	64
4.3 Peptidomics of a fish hunting cone snail	65
Comparative proteomic study of the venom of the piscivorous cone snail <i>Conus consors</i>	67
1. Introduction.....	69
2. Materials and methods.....	71
3. Results and discussion	74
References	86
4.4 Concluding remarks.....	92
4.5 Reference List	93
5. Transcriptomics	97
5.1 Introduction.....	97
5.2 The beginning: cDNA cloning	98
5.3 Next-generation sequencing (NGS)	99
High-resolution picture of a venom gland transcriptome: case study with the marine snail <i>Conus consors</i>	101
1. Introduction.....	103

2. Material and methods	104
3. Results	108
4. Discussion and conclusion.....	126
References	129
5.4 Concluding remarks.....	137
5.5 Reference List	138
6. Peptidomics and Proteomics	141
6.1 Introduction.....	141
6.2 Bioinformatics for transcriptome.....	141
6.3 Bioinformatics for peptidome and proteome.....	144
6.4 Transcriptomics to speed up proteomics and peptidomics	145
Large-scale discovery of conopeptides and conoproteins in the injected venom of a fish-hunting cone snail using a combination of proteomics and transcriptomics.....	146
1. Introduction.....	148
2. Materials and methods.....	150
3. Results and discussion	153
References	170
6.5 Concluding remarks.....	176
<i>Conus consors</i> snail venom proteomics proposes functions, pathways, and novel families involved in its venomous system.....	177
1. Introduction.....	179
2. Materials and Methods.....	182
3. Results and Discussion.....	187
4. Conclusions	208
References	210
6.6 Concluding remarks.....	217
Uncovering intense protein diversification in a cone snail venom gland using an integrative venomomics approach	218

1. Introduction.....	220
2. Materials and methods.....	223
3. Results and discussion	228
4. Conclusion.....	241
References	243
6.7 Concluding remarks.....	251
6.8 Reference List	252
7. Concluding remarks and perspectives	255
7.1 Venom peptide and protein content: towards fully integrated venomics approach.	255
7.2 Biological mechanisms in the venom duct.	255
7.3 Applications: pharmacological tools and drug candidates	257
7.4 Perspectives for venomics and drug candidate discovery	258
7.4.1 LC and MS instruments	258
7.4.2 Improving PTM detection with new algorithms	259
7.4.3 Towards the direct implementation of new bio-assay technologies in the venomics strategy.....	260
7.5 Reference list.....	261
Acknowledgements	263

List of Figures

Figure 1: Workflow of the studies presented in this thesis work.	31
Figure 2: General anatomy of the cone snail	36
Figure 3: Diagram showing the main features used for shell description.	37
Figure 4: Shell shapes vary quite a lot, from broadly conical (A), conical (B) to biconical (C).	37
Figure 5: Graphical view of the cone snail venom apparatus	39
Figure 6: Optical microscopy image of a cone snail dissection.	40
Figure 7: Main characteristics of a cone snail radula	40
Figure 8: Optical microscopy images of radula details. Black arrows show by transparency the radula's lumen with air bubbles trapped in water	41
Figure 9: Examples of three types of radulae that correspond to the different feeding habits of cone snails.	44
Figure 10: Typical disposition of the capsules containing the eggs laid by the female <i>Conus consors</i>	45
Figure 11: Larva of <i>Conus consors</i> at veliger stage, soon after hatching.	46
Figure 12: Conopeptide production process.	52
Figure 13: Amino acid conservation level comparison.	54
Figure 14: NMR structure of the CcTx conopeptide with O-glycosylation PTM.	57
Figure 15: Distribution of <i>Conus consors</i> dissected venom fractions molecular masses obtained by ESI-MS and MALDI-MS.	77
Figure 16: Comparison of <i>Conus consors</i> dissected venom fractions between MALDI-MS and ESI-MS techniques according to mass range and hydrophobicity.....	78
Figure 17: UV chromatograms at 214 nm obtained by RP-HPLC fractionation of milked venom (top) and dissected venom (bottom) of the cone snail, <i>Conus consors</i>	80
Figure 18: EST processing and description of the venom duct transcriptome contigs obtained by similarity searches.....	109

Figure 19: Gene ontology-sorted annotations.....	110
Figure 20: Recovery between 454 transcriptomic identified isoforms and previously identified sequences.....	112
Figure 21: Transcription profiling of the conopeptides cocktail.....	116
Figure 22: Alignment of A superfamily conopeptides (non CcTx).....	117
Figure 23: Phylogenetical analysis of <i>C. consors</i> A superfamily non-CcTx isoforms.....	118
Figure 24: Alignment of S superfamily toxins.....	119
Figure 25: Multiple alignment of Conopressin toxins.....	120
Figure 26: Summary of classical transcriptome analysis operations.	142
Figure 27: Improved workflow for transcriptome analysis.	143
Figure 28: Scheme of the workflow used for the peptide or protein identification.	145
Figure 29: Typical UV (214 nm) chromatogram of pooled milked venom of <i>C. consors</i>	154
Figure 30: (A) Molecular mass range distribution of the 419 native compounds detected by ESI-MS analysis. (B) Number of disulphide bridges assigned for the 209 components for which a native and reduced mass could be determined without ambiguity.	155
Figure 31: α -Conotoxin contigs from which 14 mature sequences were identified in the injectable venom.	157
Figure 32: Example of M-superfamily contig alignment with propeptide and mature toxins identified in the injectable venom.....	162
Figure 33: Alignment of linear sequence from <i>C. litteratus</i> (Q2HZ30) and <i>C. consors</i>	163
Figure 34: Alignment of actinoporin-like sequences from <i>M. echo</i> (Q76CA2) and <i>C. consors</i> (Conoporin-Cn1).	165
Figure 35: Family distribution of identified conopeptides and conoproteins in <i>C. consors</i> injectable venom.	167
Figure 36: Venom components are partially or completely secreted into the lumen of the venom duct in the form of granules.....	180
Figure 37: Venom composition of the fish-hunting cone snail <i>C. consors</i> is less complex than that of the worm-hunting cone snail <i>C. quercinus</i>	188
Figure 38: Exploring the proteomes of <i>C. consors</i> IV and DV	189

Figure 39: Evolutionary relationships within the mollusk-specific conoprotein 4 orthologous family.	205
Figure 40: UV chromatogram at 214 nm obtained by RP-HPLC fractionation of the DV of <i>Conus consors</i>	229
Figure 41: Molecular mass range distributions of native components found in the IV vs. DV.	232
Figure 42: Alignment (A) and tree (B) of all reported and new conomarphins.	235
Figure 43: Overview of the cone snail general anatomy (A) and the venom apparatus from the elaboration of the venom and venom granules in the venom gland tissues to delivery to prey (B). Family distribution of conopeptides, conoproteins and other components in the transcriptome, the DV and the IV (C).	240
Figure 44: Simplified scheme of the venom apparatus.	256

List of Tables

Table 1: Approved drugs derived from venom peptides or proteins	28
Table 2: List of the most common post-translational modifications observed in conopeptides ^{9,10}	56
Table 3: Summary of the 12 conopeptide pharmacological families described to date.....	58
Table 4: Previously described <i>Conus consors</i> venom sequences, their calculated monoisotopic masses and matching masses measured by ESI-MS in the dissected venom (DV) and milked venom (MV).....	83
Table 5: Main features of <i>Conus consors</i> conopeptides.	113
Table 6: Main features of toxin-related transcripts.	123
Table 7: A-conotoxins.....	156
Table 8 : M-conotoxins.	158
Table 9 : O1-conotoxins.	159
Table 10: A-propeptides.	160
Table 11: M-propeptides.....	161
Table 12 : Other linear conopeptides.....	163
Table 13: HMM proteins identified in the <i>C. consors</i> IV	191
Table 14: HMM proteins identified in the <i>C. consors</i> DV	192
Table 15: HMM proteins of the <i>C. consors</i> venom listed by molecular function	198
Table 16: Distribution of Conoproteins 1–5 in genome and transcriptome databases.....	204
Table 17: Difference in expression levels of genes encoding HMM venom proteins between venom duct and salivary gland.....	207
Table 18: Example of C- and N-terminus hyper-variability with DV conopeptides of the T-superfamily.....	237

Abbreviations

BLAST	basic local alignment search tool
bp	base pair
cDNA	complementary deoxyribonucleic acid
DB	database
DV	dissected venom
ESI	electrospray ionisation
EST	expressed sequence tag
GO	gene ontology
HMM	high molecular mass
HPLC	high performance liquid chromatography
HTS	high throughput screening
IEF	isoelectric focusing
IPG	immobilized pH gradient
IV	injected or injectable venom
JTT	matrix-based substitution model from Jones DT, Taylor WR and Thornton JM
LC	liquid chromatography
LC-MS	liquid chromatography coupled to mass spectrometry
LTR	long terminal repeat
MALDI	matrix-assisted laser desorption/ionisation
mRNA	messenger ribonucleic acid
MS	mass spectrometry (or mass spectrometer)
MS/MS	tandem mass spectrometry
MV	milked venom (same as IV)
NGS	next-generation sequencing
NJ	neighbour joining

Abbreviations

PCR	polymerase chain reaction
PTM	post-translational modification
QTOF	quadrupole-time-of-flight
R&D	research and development
RP-HPLC	reversed-phase high performance liquid chromatography
RT	retention time
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TE,	transposable element
TOF	time-of-flight
TRG	toxin related gene
UHPLC	ultra-high pressure liquid chromatography
UV	ultraviolet
XB	xenobiotic

Publications

Articles

1. Biass D, Violette A, Hulo N, Lisacek F, Favreau P, Stöcklin R. **Uncovering intense protein diversification in a cone snail venom gland using an integrative venomomics approach.** J Proteome Res. 2015 Feb 6;14(2):628-38. doi:10.1021/pr500583u. Epub 2015 Jan 22. PMID: 25536169.
2. Peigneur S, Paolini-Bertrand M, Gaertner H, Biass D, Violette A, Stöcklin R, Favreau P, Tytgat J, Hartley O. **δ-Conotoxins synthesized using an acid-cleavable solubility tag approach reveal key structural determinants for NaV subtype selectivity.** J Biol Chem. 2014 Dec 19;289(51):35341-50. doi:10.1074/jbc.M114.610436. Epub 2014 Oct 28. PMID: 25352593.
3. Petrel C, Hocking HG, Reynaud M, Upert G, Favreau P, Biass D, Paolini-Bertrand M, Peigneur S, Tytgat J, Gilles N, Hartley O, Boelens R, Stocklin R, Servent D. **Identification, structural and pharmacological characterization of τ-CnVA, a conopeptide that selectively interacts with somatostatin sst3 receptor.** Biochem Pharmacol. 2013 Jun 1;85(11):1663-71. doi: 10.1016/j.bcp.2013.03.019. Epub 2013 Apr 5. PMID: 23567999.
4. Leonardi A, Biass D, Kordiš D, Stöcklin R, Favreau P, Križaj I. **Conus consors snail venom proteomics proposes functions, pathways, and novel families involved in its venom system.** J Proteome Res. 2012 Oct 5;11(10):5046-58. doi:10.1021/pr3006155. Epub 2012 Sep 12. PMID: 22928724.
5. Violette A, Biass D, Dutertre S, Koua D, Piquemal D, Pierrat F, Stöcklin R, Favreau P. **Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach.** J Proteomics. 2012 Sep 18;75(17):5215-25. doi: 10.1016/j.jprot.2012.06.001. Epub 2012 Jun 13. PMID: 22705119.

6. Eugster PJ, Biass D, Guillaume D, Favreau P, Stöcklin R, Wolfender JL. **Peak capacity optimisation for high resolution peptide profiling in complex mixtures by liquid chromatography coupled to time-of-flight mass spectrometry: application to the *Conus consors* cone snail venom.** J Chromatogr A. 2012 Oct 12;1259:187-99. doi: 10.1016/j.chroma.2012.05.033. Epub 2012 May 14. PMID: 22658136.

7. Violette A, Leonardi A, Piquemal D, Terrat Y, Biass D, Dutertre S, Noguier F, Ducancel F, Stöcklin R, Križaj I, Favreau P. **Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of *Conus* venoms.** Mar Drugs. 2012 Feb;10(2):258-80. doi: 10.3390/md10020258. Epub 2012 Jan 31. PMID: 22412800.

8. Terrat Y, Biass D, Dutertre S, Favreau P, Remm M, Stöcklin R, Piquemal D, Ducancel F. **High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*.** Toxicon. 2012 Jan;59(1):34-46. doi: 10.1016/j.toxicon.2011.10.001. Epub 2011 Nov 6. PMID: 22079299.

9. Dutertre S, Biass D, Stöcklin R, Favreau P. **Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity.** Toxicon. 2010 Jul;55(8):1453-62. doi: 10.1016/j.toxicon.2010.02.025. Epub 2010 Mar 3. PMID: 20206197.

10. Biass D, Dutertre S, Gerbault A, Menou JL, Offord R, Favreau P, Stöcklin R. **Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*.** J Proteomics. 2009 Mar 6;72(2):210-8. doi: 10.1016/j.jprot.2009.01.019. Epub 2009 Jan 23. PMID: 19457347.

Presentation and posters

Fall Meeting of the Swiss Chemical Society (SCS) in Lausanne, Switzerland, September 2011, Poster

Optimization of *Conus consors* venom profiling using Ultra-High Pressure Liquid Chromatography (UHPLC).

Eugster PJ, Biass D, Guillaume D, Favreau P, Stöcklin R and Wolfender JL.

18th Meeting of the “Société Française pour l’Etude des Toxines” (SFET), 2010, Oral presentation

Rapid intraspecimen evolution of cone snail venom composition: multiplying pharmacological profiles?

Biaass D, Dutertre S, Favreau P and Stöcklin R.

21st American Peptide Symposium (APS), 2009, Poster

Peptidomics and proteomics of *Conus consors* cone snail venom

Biaass D, Križaj I, Leonardi A, Dutertre S, Favreau P and Stöcklin R.

4th International congress of Natural Peptide to Drug (NP2D), 2009, Oral presentation

Peptidomics of *Conus consors*

Biaass D, Favreau P and Stöcklin R.

1. Venomics

1.1 Introduction

The recent tendency to group thematically-related fields and techniques of Life Sciences around a theme word ending in “omics” has naturally led the venom specialists to create the term “Venomics” for their purposes. The latter was introduced in the early 2000s when studies on snake venom protein contents took a more comprehensive approach^{1,2}. Venomics covers techniques related to the study of venoms, such as proteomics, peptidomics and transcriptomics³. The combination of different methods and techniques into one single approach is essential not only to understand the complex and specific underpinnings of a given scientific domain, but also to grasp the global picture that the connections between these domains create. In a similar way that other “omics” were ushered into service with a view to extracting the most pertinent knowledge in applied sciences or fundamental research, venomics was not only introduced to solely characterise new compounds for their use as novel pharmacological research tools or new drug lead candidates. Indeed, venomics’ other virtue is that it almost de facto triggers an instinct to understand both nature’s “thinking pattern” in its necessity to endow certain animals with such powerful means of prey capture or defence, and in what way this process could be beneficial to humans.

1.2 Venomous animals and their venoms

Why study venomous animals and their venoms? How can they be beneficial to humans in spite of the negative aspects they convey?

Venomous species – more than 100,000 – are spread throughout the world. Not only are they represented in main groups of terrestrial animals like arachnids, insects, snakes, frogs, etc, but they also feature in many major groups of marine life families such as corals, sponges, sea anemones, cone snails, etc ⁴. For the scientific community, venoms have been of high interest for many, many years. Can the high potency of a venom, to guarantee the survival of a venomous animal, always find a true justification? Several decades of scientific studies have shown that venoms are efficient combinations of hundreds of bioactive substances of various nature: alkaloids, aliphatic acids, amines, histamines, steroids, peptides, proteins and many others ⁵. On the other hand, venomous animal envenoming potential results from the vast array of available exogenous targets to be “treated”: neurone and neuromuscular junctions, enzymes, ion channels, receptors and their possible modulators, etc. Not to mention the corollary, i.e. that these targets will be affected differently according to the potency, affinity and selectivity of the venom substances⁵. Given the hundreds of molecules constituting a single venom and the number of venomous animal species, nature provides an incredible reservoir of bioactive components⁶. Unfortunately, only a small percentage of molecules have been described so far from a structural and functional point of view ⁷.

1.3 From natural products to drug candidates

Venom compounds are natural products. However, unlike other natural products like plant molecules, the recognition of their real potential as research tools and drug candidates in the medicine world occurred much later, something that is certainly due to their structural complexity. Indeed, small natural molecules have occupied a major position in medicine and drug discovery for thousands of years. Unfortunately, this interest in natural products greatly decreased in the early 90s. Important developments in pharmaceutical research such as high

throughput screening (HTS) have pushed pharmaceutical industries to change their gaze and focus on new fields such as combinatorial chemistry, which have saturated the research and development market with a considerable amount of compound libraries^{8,9}. Those developments were supposed to reduce costs and provide much faster results for drug discovery than the use of natural products and/or traditional methods. As a consequence, investments in natural products decreased because the techniques to develop a natural product into a drug candidate were deemed too slow and expensive. In addition, the propensity of natural products to render intellectual property issues slightly more difficult to tackle combined with new conventions on biodiversity, like the 1992 Rio convention, hemmed in the enthusiasm of pharmaceutical companies to use biomaterial collections for their research.

However, even though great technological advances were made in many critical fields for pharmaceutical research and development, the number of approved drugs arriving on the market has continuously decreased¹⁰ and could explain why a renewal of interest in natural products and natural product-derived compounds as a source for new lead candidates in drug discovery has recently taken place.

Indeed, neither HTS nor combinatorial chemistry have held their promises to replenish the drug development pipelines. In spite of the tremendous amount of new combinatorial chemistry compound libraries that have been created and tested, only one drug candidate developed exclusively from *de novo* combinatorial chemistry has made it through to approval until now: Sorafenib, an antitumor molecule co-developed by Bayer and Onyx Pharmaceuticals as Nexavar¹¹. This could be explained by the intrinsic synthetic structure of combinatorial chemistry components to give them, surprisingly, a narrower diversity than natural products¹².

Other factors have been instrumental in the renewal of interest in natural compounds. The very fast development, speed acceleration and accuracy of proteomic and bioinformatic

techniques, particularly of protein separation and bioinformatic-guided structure elucidation, have contributed to overcome certain issues in the study of natural products structures and structure-function relations¹³. Not to be left aside either, is the new product marketing trend in favour of ecological uses of what nature's biodiversity can offer to human health.

1.4 What about venom compounds as drug candidates?

In the light of the renewed interest for natural products explained above, venom components constitute a major asset for the future of new drug research and development effort. As explained earlier, venom components are quite attractive from a pharmacological point of view as their targets for envenomation are very close to those of certain pathologies, including chronic pain, cancers, HIV, autoimmune diseases and many neuronal diseases¹⁴. So far, pharmaceutical research and development in venom compounds have actually led to the approval of seven venom-derived drugs (see Table 1 modified from Ref. ¹⁴).

Table 1: Approved drugs derived from venom peptides or proteins

Peptide or protein (or derivative)	Animal source	Molecular target	Indication	FDA approval year
Captopril	Pit viper	Angiotensin-converting enzyme (ACE)	Hypertension	1981
Eptifibatide	Pygmy rattlesnake	IIb/IIIa integrin receptor	Acute coronary syndromes	1998
Tirofiban	Saw-scaled viper	IIb/IIIa integrin receptor	Acute coronary syndromes	1999
Bivalirudin	Medicinal leech	Thrombin	Coagulation during surgery	2000
Ziconotide	Magician cone snail	Ca _v 2.2 channel	Chronic pain	2004
Exenatide	Gila monster lizard	GLP-1 receptor	Type 2 diabetes	2005
Batroxobin	Lancehead snake	Fibrinogen	Perioperative bleeding	Outside USA

Several other venom-derived compounds are in the pipelines of certain pharmaceutical companies: 9 are in clinical trial phases and 10 in preclinical studies¹⁴. Many of these drugs, that are either on the market or in preclinical or clinical phases, have originated from snake venoms. This could be explained by the choice of well studied targets available then, mainly in the cardiovascular system. Nowadays, many new pharmacological targets – voltage-gated channels, acid sensing ion channels, G-protein-coupled receptors, etc – are available to the pharmaceutical research, thanks to new developments in the fundamental study of the nervous system for example. These new pharmacological targets seem to fit quite nicely those of the venoms of animals such as spiders, scorpions and cone snails.

1.5 The CONCO project

In this context, and with a view to boosting the discovery of natural compound-based drugs, a new project, integrated to the “Venomics” programme initiated by the International Society on Toxinology (IST)³, was launched: CONCO, the cone snail genome project for health. The main goal of this project was to discover new molecules originating from marine venomous animals: cone snails. Thanks to applied cone snail venomomics, accelerated, cheaper, safer and more ethical production of innovative biomedical drugs could be achieved. For this purpose, a piscivorous cone snail species was chosen, the *Conus consors*. The CONCO project was funded by the sixth research framework programme (FP6) of the Research and Technological Development branch of the European Commission for a budget of 15'619'954 Euros. It involved 20 partners – universities, private companies, institutes or foundations – from 13 countries. Each partner took charge of different tasks defined by the project: administration, collection missions, sampling, biodiversity, peptidomics, proteomics, genomics, transcriptomics, bioinformatics, bioactivity assays and communication. The CONCO project also settled partnerships for the collection missions with the “Institut de

Recherche pour le Développement (IRD)” in New-Caledonia (French territory), the “Museum National d’Histoire Naturelle (MNHN)” in France and the local governments of New Caledonia and French Polynesia.

1.6 Thesis work

As the initiator of, and one of the main partners in, the project, Atheris Laboratories was entrusted with, inter alia, the peptidomic, proteomic, transcriptomic and bioinformatic portion of the research work. The following studies were achieved as part of this task.

The starting point of this task entailed the development of new methods to not only extract a maximum of sequence information from cone snail venoms, but also to better understand the various steps that take place in the venom fabrication and the envenomation processes. Following a number of preliminary tests on some cone snail species, the piscivorous *Conus consors* was chosen for the definitive studies. Regarding proteomics or peptidomics alone, the first goal was to determine what were the weaknesses of the methods used thus far; why have only a small percentage of venom compounds been described so far? A number of scenarios will be discussed here, including the sampling methods and of course the different analytical methods. These have in turn led to improvements to existing techniques applied in proteomics and their application to venomics. In order to reach the final goal – the faster and more efficient generation of biological results – not only were new methods used, but also new molecular biology techniques introduced in venomics. The process of combining the two worlds of proteomics and molecular biology could not have been achieved without the help of yet another brainchild of science, namely bioinformatics. The latter was instrumental for the funnelling of all the generated data and the final extraction of the required valuable sequence information. The workflow of the studies described in this paper is summarised in Figure 1.

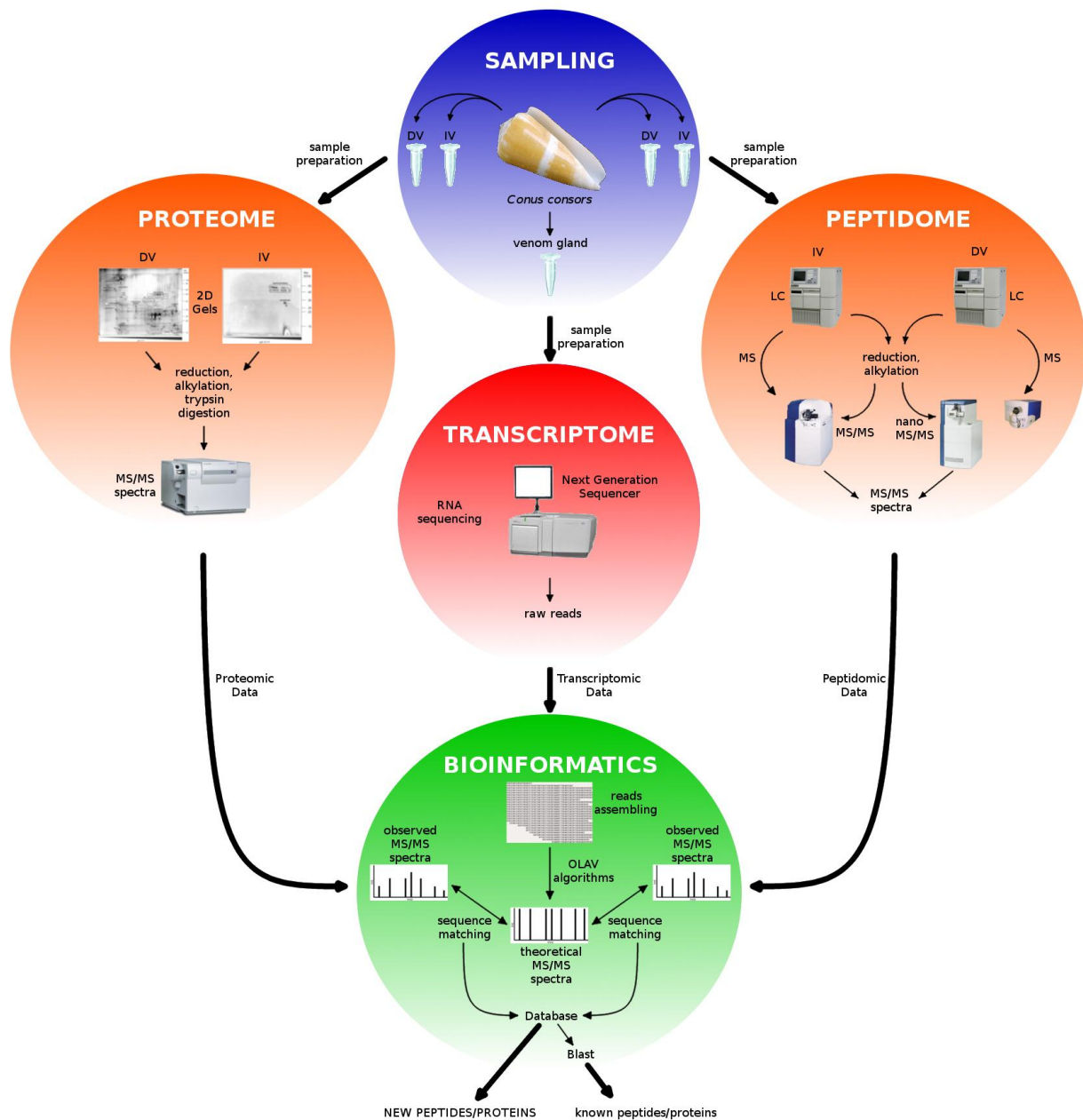


Figure 1: Workflow of the studies presented in this thesis work.

Following different sampling methods (top), resultant samples are submitted to proteomic, peptidomic and transcriptomic techniques (middle) to finally obtain valuable sequence information through bioinformatics (bottom).

This thesis was involved in the following areas of the flowchart given above:

- most aspects of the sampling (all species),

- preliminary studies on the *Conus quercinus* venom (not presented here).
- part of the data analysis of the *Conus consors* transcriptome,
- all the peptide mass matching of the *Conus consors* injected venom (IV) and the dissected venom (DV) proteome (bioinformatics) and part of the data analysis of the IV and DV proteomes,
- part of the data analysis of the IV peptidome,
- all the work achieved on the DV peptidome.

Some transcriptomic sequencing of other cone snail species was achieved, but not presented here, with the exception of the analysis of some sequence data for the final peptidomic study presented here.

1.7 Reference list

1. Juarez, P.; Sanz, L.; Calvete, J. J. Snake venomomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics*. 2004, 4 (2), 327-338.
2. Bazaa, A.; Marrakchi, N.; El, A. M.; Sanz, L.; Calvete, J. J. Snake venomomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*. *Proteomics*. 2005, 5 (16), 4223-4235.
3. Menez, A.; Stocklin, R.; Mebs, D. 'Venomics' or : The venomous systems genome project. *Toxicon* 2006, 47 (3), 255-259.
4. Mebs, D. *Venomous and Poisonous Animals*; Medpharm ed.; 2002.
5. Meier, J. Venomous and Poisonous Animals - A Biologist's View. In *Handbook of Clinical Toxicology of Animal Venoms and Poisons*, CRC Press ed.; Meier, J., White, J., Eds.; 1995.
6. Escoubas, P.; King, G. F. Venomics as a drug discovery platform. *Expert. Rev. Proteomics*. 2009, 6 (3), 221-224.
7. Stocklin, R.; Favreau, P. Proteomics of venom peptides. In *Perspectives in molecular toxinology*, - ed.; Menez, A., Ed.; John Wiley & Sons Ltd: Chichester, UK, 2002; pp 107-123.
8. Ortholand, J. Y.; Ganesan, A. Natural products and combinatorial chemistry: back to the future. *Curr. Opin. Chem. Biol.* 2004, 8 (3), 271-280.
9. Koehn, F. E.; Carter, G. T. Rediscovering natural products as a source of new drugs. *Discov. Med.* 2005, 5 (26), 159-164.
10. Pammolli, F.; Magazzini, L.; Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat Rev. Drug Discov.* 2011, 10 (6), 428-438.

11. Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat Prod.* 2012, 75 (3), 311-335.
12. Feher, M.; Schmidt, J. M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 2003, 43 (1), 218-227.
13. Fox, J. W.; Shannon, J. D.; Stefansson, B.; Kamiguti, A. S.; Theakston, R. D. G.; Serrano, S. M. T.; Camargo, A. C. M.; Sherman, N. Role of Discovery Science in Toxinology : Examples in Venom Proteomics. In *Perspectives in Molecular Toxinology*, - ed.; Menez, A., Ed.; John Wiley & Sons: 2002.
14. King, G. F. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert. Opin. Biol. Ther.* 2011, 11 (11), 1469-1484.

2. Cone snails

2.1 Introduction

Cone snails are part of the broad phylum of Mollusca and the very large superfamily of Conoidea, also known as Toxoglossa, which literally means “poison tongue”.

The different families of this group, Terebridae, Turridae (subdivided in 13 new monophyletic families ¹) and Conidae, have the peculiarity of possessing a sophisticated venom apparatus for the capture of their prey, and are thus considered as a monophyletic group. Different studies still argue on the total number of species in this superfamily, but it is estimated that it contains about 10,000 species ².

The Conidae family includes approximately 700 species commonly referred to as cone snails, cones or cone shells by the conchologists.

Their vernacular name very simply stems from the morphology of their shell that resembles more or less the geometric cone shape.

2.2 Biology of the cone snail

2.2.1 General morphology

Although from a morphological point of view Gastropoda may significantly vary from one group to another, they all present a number of common features such as one or two pairs of sensory tentacles bearing the eyes and a muscular ventral foot, hence their name which is derived from the Greek *gaster*, stomach, and *poda*, foot.

2. Cone snails

Cone snails have further characteristics common to Gastropoda, such as a torsion (occurs during a development process), a shell and a mouth, but more importantly apomorphic characteristics like a siphon, a proboscis, a radular sac, a venom gland and a muscular bulb (Figure 2).

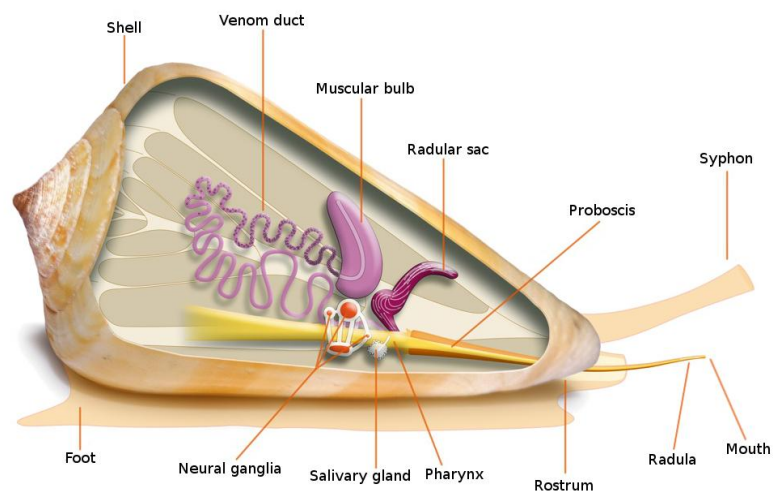


Figure 2: General anatomy of the cone snail

Traditionally, cone snail species identification was based on their shell features. These features may be reasonably easy to identify like for example the presence or not of a sculptured shell shoulder, but in many cases, they are highly variable like the shell length or width, which therefore complicates identification.

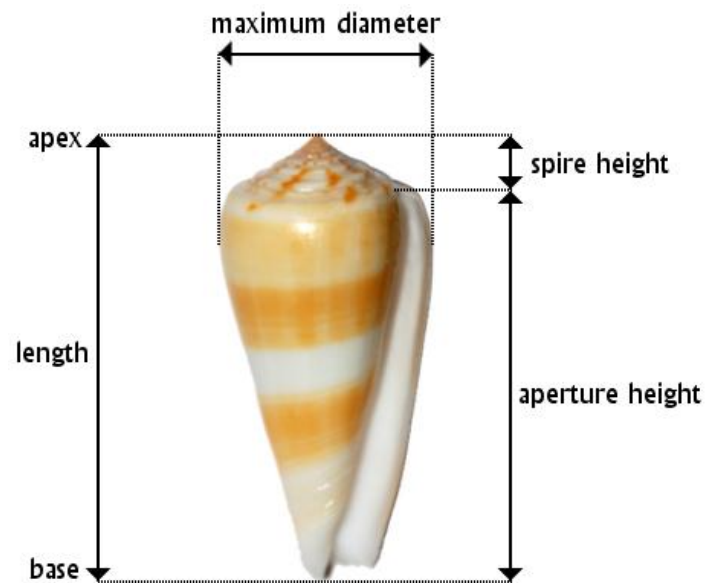


Figure 3: Diagram showing the main features used for shell description.

The main features of a shell are determined by length, maximum diameter and aperture height (Figure 3). Dozens of other features are used by conchologists to thoroughly describe cone shell or shells in general 3. Some aspects of the shell, like the shape, colours and patterns are also instrumental for proper identification (Figure 4).

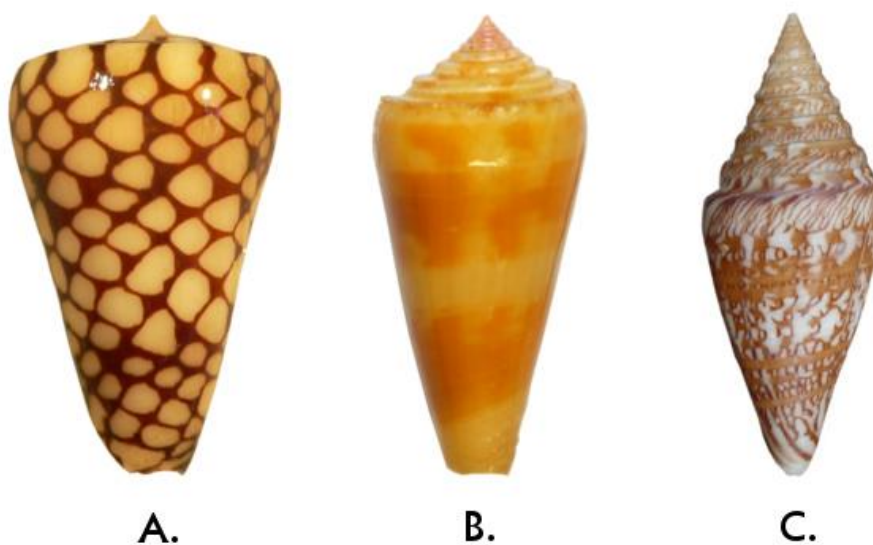


Figure 4: Shell shapes vary quite a lot, from broadly conical (A), conical (B) to biconical (C).

Differences in patterns and colours are also very useful for correct identification.

2.2.2 Venom apparatus

As mentioned previously, the main apomorphic characteristic of the cone snail is the presence of a sophisticated venom apparatus. Although the radula or single tooth is an attribute that has evolved from the small teeth common to other Gastropoda ⁴, the development of a venom gland and thus a venom associated to this radula is a unique feature of the Conoidea superfamily. The main application of this system and its advantage over the Gastropoda's primary configuration are alimentary. Indeed, it enables the cone snail to hunt and feed on more complex preys, such as fish, that would otherwise remain impossible preys for such a small and slow animal. In a second role, it can be used for self defence.

The venom apparatus consists of seven main components (Figure 5):

- the muscular bulb or venom bulb, sort of oblong muscle used to pump the venom out of the venom gland like a peristaltic pump;
- the venom duct, a narrow flexible tube (the length can vary considerably according to the species) in which the venom is produced, matured and then secreted in its lumen;
- the radular sac, kind of quiver where radulae are produced and stocked, waiting for use;
- the radulae, like harpoons or arrows of different shapes, length and width according to the feeding habits of the cone snail. They have a small ligament at their base to enable the cone snail to hold them back with their proboscis;
- the pharynx, sort of crossroads where the venom gland connects to the radular sac and meets the proboscis. It may also play a role in the fine selection mechanism for the final venom composition;
- the salivary glands, possibly used to dilute the venom. They also appear to be responsible for the production of certain toxins;

- the proboscis (which is a highly extensible muscle – akin to an elephant trunk – and to the tip of which a readily armed radula is affixed) that the cone snail can stretch out to over its own length to catch its prey and draw it back to its mouth.

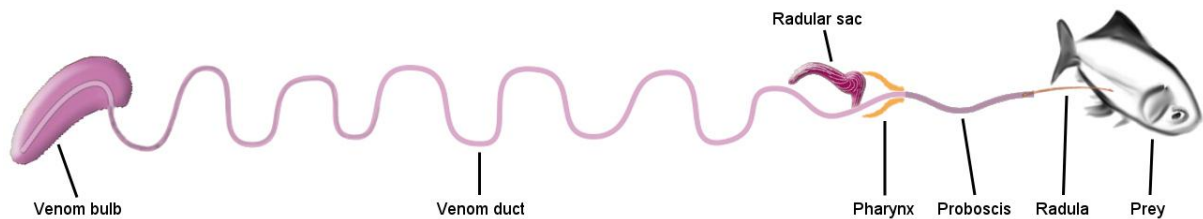


Figure 5: Graphical view of the cone snail venom apparatus

One of the first descriptions of a cone snail's anatomy dates back to 1914, when Shaw provided a highly detailed anatomical view of the venom apparatus ⁵. The venom bulb and venom duct were at the time respectively called “poison gland” and “poison gland duct”.

2.2.2.1 The venom bulb

Despite its appearance (Figure 5), the venom bulb is not the venom producing organ, as it is mainly composed of muscular tissue. Indeed, the proteome study of the venom bulb has shown that the proteins present in this organ could be split into three main functional groups: cytoskeletal, motor activity and energy metabolism proteins ⁶. High levels of specific proteins such as arginine kinase and myosin clearly demonstrate a predominant role of the venom bulb in muscular movement. They enable a rapid contraction that loads the venom into the venom delivery apparatus - the radula - for subsequent injection into the prey.

2.2.2.2 The venom duct

For the toxinology field, the venom duct is the most studied organ in the cone snail as it is the main venom producing organ. The venom duct is a relatively long organ: in some species its

unfolded length can be well over twice that of the cone shell. In vivo, the organ is folded and convoluted but the proximal side of it always goes through the nerve ring, itself composed of nervous ganglions.



Figure 6: Optical microscopy image of a cone snail dissection.

The detail shows the venom duct neatly packed together. By transparency, the venom shows white to yellow colours.

2.2.2.3 The radulae and the radular sac

The radula is one of the cone snail's most important assets for hunting. It is shaped from a rolled sheet of chitin. Each radula is formed and stocked in the radular sac, awaiting usage.

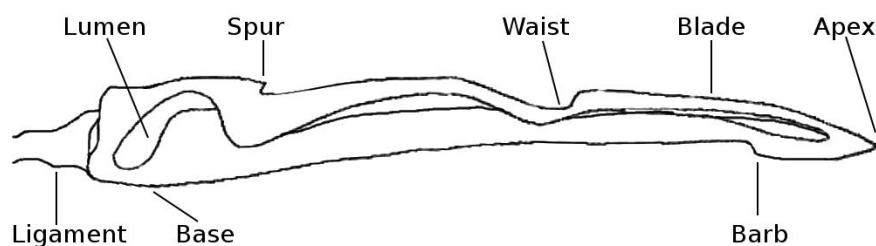


Figure 7: Main characteristics of a cone snail radula

While the shape and characteristics of the radula may vary according to species (see 2.2.3 Habitats and feeding habits), its operating elements remains unchanged (Figure 7):

- the ligament enables the cone snail to attach the radula to its base and draw the attacked prey back to its mouth. It is possibly a rudiment of the radular membrane (radulae production site in other Conoidea).
- the lumen through which the venom enters, coming from the venom duct and proboscis, and can flow to the apex of the radula (like a hypodermic needle) for the effective injection in the flesh of the prey (Figure 8).

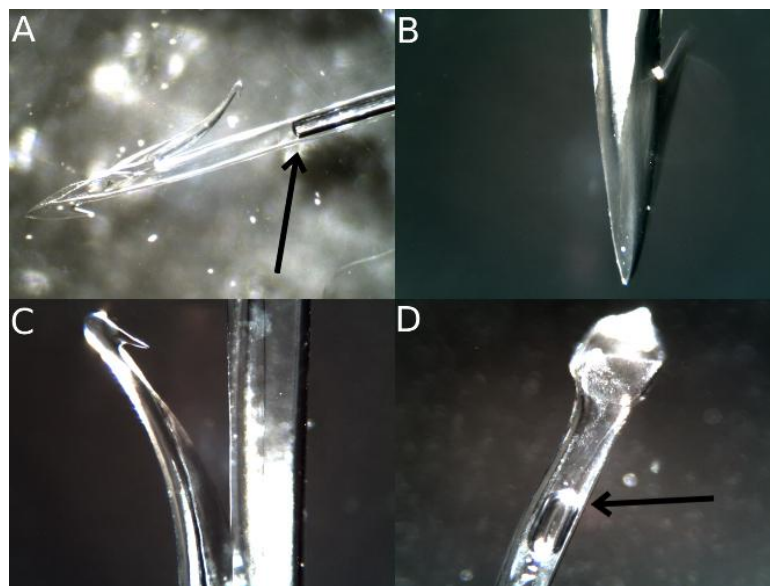


Figure 8: Optical microscopy images of radula details. Black arrows show by transparency the radula's lumen with air bubbles trapped in water

A) Apex side of a *Conus consors* radula showing the short blade but long barb characteristic of a piscivorous cone snail. **B)** Apex of a *Conus striatus*. **C)** Detail of the very impressive barb of the piscivorous *Conus striatus* radula. **D)** Base of a *Conus textile* radula.

- the blade and the barb are the two features that may mostly vary according to the feeding habits. Short blade and long barb are typical of piscivorous cone snail,

whereas long blade and small barb are traits of vermivorous cone snails (see 2.2.3 Habitats and feeding habits).

Radulae are produced and stocked by the dozen in the radular sac ready for use. The exact mechanism of fabrication has not been extensively studied in *Coninae*. The radulae are supposedly produced by odontoblast cells in a specific region of the radular sac, but unlike the other conoideans, cone snails totally lack radular membrane ⁷.

2.2.2.4 The pharynx or buccal cavity

The pharynx is at the base of the proboscis where the radular sac connects together with the oesophagus, the salivary glands and the venom duct ⁷. It is surrounded by a thick circular muscle and is close to the nerve ring. It may also play a key role in the envenomation mechanism.

2.2.2.5 The salivary glands

The salivary glands, present in all *Coninae* generally in pairs, are acinous and directly connected to the pharynx ⁷. The secretory function of the salivary glands was established by histochemical techniques ⁸, but the actual content secreted has not been extensively studied. It was suggested though that salivary gland specific conotoxins may be produced, as alpha-conotoxin-like transcripts have been found in the sequence analysis of salivary gland tissue cDNA library ⁹.

2.2.2.6 The proboscis

The proboscis is a very important prey capture element. The organ is tube shaped and can reach out to twice the length of the cone shell. The distal tip of the proboscis is the true mouth. This is where the radula is held thanks to the buccal tube sphincter ⁷. Once the prey

is stabbed, the cone snail can retract its proboscis in its rostrum. Recently, an extensive study of the proboscis tip demonstrated that it also plays a very important role in prey location. Indeed, sensory papillae have been located by electron microscopy on the proboscis tip, clearly suggesting that the proboscis is an essential organ for prey localisation and identification (according to cone snail diet) prior to envenomation ¹⁰.

The mechanism of action of the proboscis during prey capture has also been studied. The radula is propelled into the prey at an extremely high speed - more than 3 m/s - although it still remains unclear how the proboscis generates such pressure to propel the radula into the prey ¹¹.

2.2.3 Habitats and feeding habits

With more than 700 living *Conus* species ¹², this genus has primarily evolved in tropical regions (mainly Indo-Pacific region), but has also spread in many other areas of the globe – from seashore to abyssal depths. This widespread distribution gives every reason to believe that many other species are still to be discovered.

Cone snails live in different environments, but mostly close to or on coral reefs. Some bury themselves in the sand around the reefs, allowing only their siphon to protrude for breathing, others hide in crevasses or small cavities in the reef corals and finally some will slip under small rocks situated in the intertidal zone (data collected from the different missions during the Conco project). Their different habitats are more or less related to their hunting manners and therefore feeding habits.

Three different types of feeding habits are known for cone snails: vermivorous, molluscivorous and piscivorous. These diets are highly correlated to the type of radulae that each cone snail produces. Indeed, the structure of the radula varies according to the type of prey the cone snail hunts (Figure 9). The radulae from vermivorous cone snails are in general broader and smaller with simple arrow shaped tips. Those of the molluscivorous type

are of intermediate length and thinner with slightly more convoluted tips. Finally, piscivorous cone snails radulae are long and thin, and have the most sophisticated tips with one or several barbs to make sure that the fish, more vigorous than other preys, cannot escape.

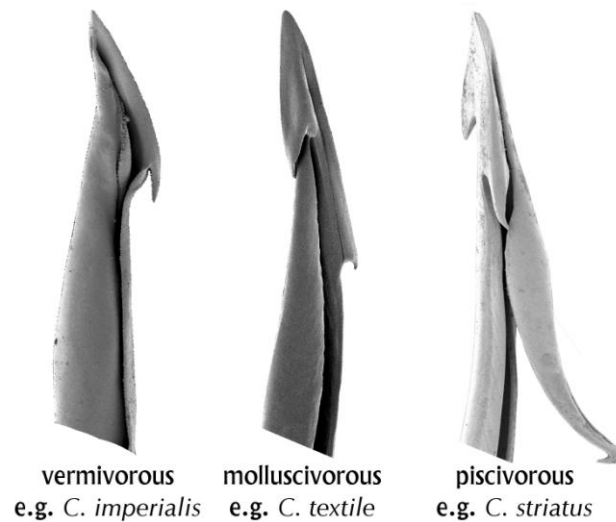


Figure 9: Examples of three types of radulae that correspond to the different feeding habits of cone snails.

Given these differences, radular morphology was also used in the species categorisation¹³. Indeed, like for the cone shell, the radula has numerous features that can be used for the classification of cone snails (Figure 7).

2.3 Reproduction and Development of the cone snail

Very little scientific research has been undertaken on the reproduction and lifecycle of cone snails in general, especially in their natural habitat. Cone snails appear to have a dioecious reproduction, meaning that individuals have separate sexes. In the context of this work, the reproduction and first stages of development have been observed with *Conus consors*

species hosted in an aquarium. After mating, the female laid several capsules, generally attached underneath rocks (Figure 10). Each capsule may contain dozens to hundreds of eggs depending on the species.



Figure 10: Typical disposition of the capsules containing the eggs laid by the female *Conus consors*.

Several dozens of eggs can be distinguished by transparency through the capsules hooked under the rock.

A few days after the capsules were laid, eggs started hatching and larvae in the veliger stage could be observed (Figure 11). These larvae were fed with specific phytoplankton (a mix of *Isochrysis galbana* and *Phaeodactylum tricornutum*) for a couple of days. Larvae grew a couple of mm, but unfortunately all died after 3 days. A couple of studies confirm that it is extremely difficult to observe the full development of cone snails in tanks with artificial water as the mortality rate in natural waters already reaches over 99% ^{14,15}.



Figure 11: Larva of *Conus consors* at veliger stage, soon after hatching.

Under binocular microscope at 100x magnification, the two lobes of the velum can distinctly be observed, as well as the torsion of the shell in the background. Cilia run all along the perimeter of each velar lobe. These are used for propulsion and food collection. During feeding, the cilia movements create a current throwing the phytoplankton directly into the mouth in the centre, between the two lobes. The two black spots in the middle are the eyes of the larva.

2.4 Evolution of the classification

Cone snails are part of the Conidae family which is the most intensively studied branch of the Conoidea superfamily. Common usage was to name each living cone snail to a single genus, *Conus*¹⁶, the traditional classification being mainly based on shell morphology. However, more recently, anatomical studies have reshuffled the groups in the Conoidea superfamily, displacing some groups from one family to another¹⁷. Another study based on the radular morphology also led to a complete revision of the Conoidea superfamily groups and subgroups¹⁸. The authors even propose to divide the Conidae family in several smaller units, renaming the genus of most cone snail species on the base of the traditional subfamily designation. Nowadays, with the increasing amount of material collected and the new molecular biology techniques, a vast amount of data is generated enabling to fine tune the description and classification of species in the Conoidea superfamily. The most recent attempt to revise the classification of cone snails by molecular phylogeny finally divided the Conidae family into four genera, *Conus* (which contains more than 80% of the *Conidae* species), *Californiconus*, *Conasprella* and *Profundiconus*, each of them being then subdivided into subgenera, like *Pionoconus* in the case of the *Conus consors* for example¹⁹. This is in agreement with the previous study based on the radular morphology in the case of the Conidae family¹⁸

2.5 Reference list

1. Bouchet, P.; Kantor, Yu. I.; Sysoev, A.; Puillandre, N. A new operational classification of the Conoidea (Gastropoda). *Journal of Molluscan Studies* 2011, 77, 273-308.
2. Bandyopadhyay, P. K.; Stevenson, B. J.; Ownby, J. P.; Cady, M. T.; Watkins, M.; Olivera, B. M. The mitochondrial genome of *Conus textile*, *coxI-coxII* intergenic sequences and Conoidean evolution. *Mol. Phylogenet. Evol.* 2008, 46 (1), 215-223.
3. Rockel, D.; Korn, W.; Kohn, A. J. *Manual of the Living Conidae*; - ed.; Verlag Christa Hemmen: 1995; Vol. -.pp. 1-517.
4. Taylor, J. D.; Kantor, Y. I.; Sysoev, A. V. Foregut anatomy, feeding mechanisms, relationships and classification of the Conoidea (=Toxoglossa)(Gastropoda). *Bulletin of the Natural History Museum of London (Zoology)* 1993, 59 (2), 125-170.
5. Shaw, H. O. N. On the anatomy of *Conus tulipa*, Linn., and *Conus textile*, Linn. *Quarterly Journal of Microscopical Science* 1914, 60 (237), 1-U18.
6. Safavi-Hemami, H.; Young, N. D.; Williamson, N. A.; Purcell, A. W. Proteomic interrogation of venom delivery in marine cone snails: novel insights into the role of the venom bulb. *J. Proteome. Res.* 2010, 9 (11), 5610-5619.
7. Taylor, J. D.; Kantor, Y. I.; Sysoev, A. V. Foregut anatomy, feeding mechanisms, relationships and classification of the Conoidea (=Toxoglossa)(Gastropoda). *Bulletin of the Natural History Museum of London (Zoology)* 1993, 59 (2), 125-170.
8. Schultz, M. C. A correlated light and electron microscopic study of the structure and secretory activity of the accessory salivary glands of the marine gastropods, *Conus flavidus* and *C. vexillum* (neogastropoda, conacea). *J. Morphol.* 1983, 176 (1), 89-111.
9. Biggs, J. S.; Olivera, B. M.; Kantor, Y. I. Alpha-conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon* 2008, 52 (1), 101-105.

10. James, D.; Prator, C. A.; Martin, G. G.; Schulz, J. R. Morphology of sensory papillae on the feeding proboscis of cone snails (Mollusca, Gastropoda). *Invertebrate Biology* 2014.
11. Schulz, J. R.; Norton, A. G.; Gilly, W. F. The projectile tooth of a fish-hunting cone snail: *Conus catus* injects venom into fish prey using a high-speed ballistic mechanism. *Biol. Bull.* 2004, 207 (2), 77-79.
12. Bouchet, P.; Gofas, S. *Conus* Linnaeus, 1758. *World Register of Marine Species* 2013, -.
13. Tucker, J. K.; Tenorio, M. J. *Systematic Classification of Fossil and Recent Conoidean Gastropods*; - ed.; Hackenheim, 2009.
14. Perron, F. E. Laboratory Culture of the Larvae of *Conus* Textile Linne (Gastropoda, Toxoglossa). *Journal of Experimental Marine Biology and Ecology* 1980, 42 (1), 27-38.
15. Perron, F. E. Growth, Fecundity, and Mortality of *Conus-Pennaceus* in Hawaii. *Ecology* 1983, 64 (1), 53-62.
16. Rockel, D.; Korn, W.; Kohn, A. J. *Manual of the Living Conidae*; - ed.; Verlag Christa Hemmen: 1995; Vol. -.pp. 1-517.
17. Taylor, J. D.; Kantor, Y. I.; Sysoev, A. V. Foregut anatomy, feeding mechanisms, relationships and classification of the Conoidea (=Toxoglossa)(Gastropoda). *Bulletin of the Natural History Museum of London (Zoology)* 1993, 59 (2), 125-170.
18. Tucker, J. K.; Tenorio, M. J. *Systematic Classification of Fossil and Recent Conoidean Gastropods*; - ed.; Hackenheim, 2009.
19. Puillandre, N.; Bouchet, P.; Duda, T. F., Jr.; Kauferstein, S.; Kohn, A. J.; Olivera, B. M.; Watkins, M.; Meyer, C. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol. Phylogenet. Evol.* 2014, 78C, 290-303.

3. Conopeptides and Conoproteins

3.1 Introduction

Like most animal venoms, the venom of a marine cone snail is a rich cocktail of various components. As discussed later, the venom content depends on the type of cone snail and on the venom sample examined. The most common biologically active components found in a cone snail venom are peptides – mostly between 10 and 30 amino acids in length – but also seen are several medium- and large-size proteins that take part either in the maturation or in the envenomation processes itself. Because of the extensive pharmacological diversity cone snail venom peptides offer, they have been a focal point of cone snail venom exploration for a while, although even its proteins have attracted scientist's interest^{1,2}. These peptides are usually called conotoxins or conopeptides, the latter term being preferred due to its less aggressive tone and more “druggable” connotation. They boast a great variety of structures owed to different disulfide bridge arrangements, as well as many different posttranslational modifications (PTMs), which greatly contribute to their potency, efficacy and specificity against a number of differing targets^{3,4}.

3.2 Conopeptides biosynthesis and classification

A brief description of conopeptide biosynthesis is herewith provided to allow one to better grasp the classification process of such conopeptides.

Conopeptides are generated by their corresponding genes in the venom duct cells. The genes are transcribed into messenger nucleic acid precursors (mRNA), which are then in turn translated into peptide precursors. These will then finally turn into mature peptides after

maturation process, where the signal sequence and propeptide regions are cleaved, to possibly be active during envenomation on specific targets of the prey, like membrane receptors, ion channels, etc (Figure 12).

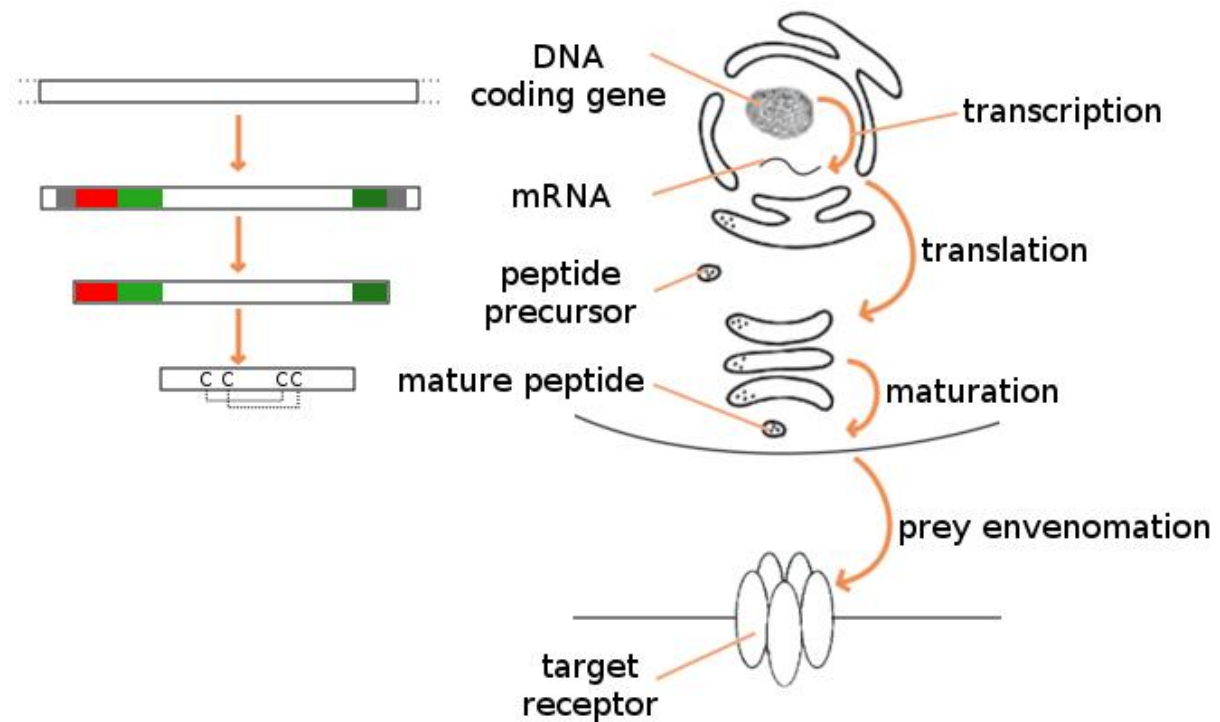


Figure 12: Conopeptide production process.

The different conopeptide classifications relate to their production and maturation process and their activity. Conopeptides can be classified into gene superfamilies, based on their signal sequence (in red) in the mRNA transcript. They can also be classified into disulfide frameworks, according to the presence or absence of cysteine residues, and their number and pattern if these are present. Finally they can be classified into pharmacological families according to their targets and their action on these targets. On the peptide precursor form, the red part corresponds to the signal sequence, the green parts to the propeptides regions and the white part to the mature peptide sequence.

The following conopeptide classification scheme has been commonly used and accepted by the conopeptide scientific community over the last decades. Conopeptides have been classified into different categories on the basis of their production and maturation processes and of their activities. The first step in conopeptides classification is to discriminate disulfide-rich (with two disulfide bridges or more) from disulfide-poor (with one or no disulfide bridge) specimens. Based on the mRNA transcript signal sequence, conopeptides can then be classified into gene superfamilies. The signal sequences are highly conserved regions in the precursor peptide in opposition to the mature sequence (see Figure 13). They can be classified into disulfide frameworks according to the cysteine residue number and arrangement in the mature peptide. Finally, they can also be classified into pharmacological families according to the biological activity of the mature peptide. Based on the most extensive conopeptides repository, i.e. the ConoServer database (www.conoserver.org), 26 conopeptide superfamilies (identified by Latin alphabet capital letters), 26 cysteine frameworks (identified by roman numbers) and 12 pharmacological families (neurological targets) (identified by Greek alphabet letters) have been to date described⁵.

While, by definition, the disulfide framework classification can only be used with disulfide-rich conopeptides, the gene superfamily classification can also be used with the disulfide-poor conopeptide. For example, the signal sequences of conomarphins (i.e. disulfide-poor conopeptides that will be discussed later in a study case) are homologous to those of the M-superfamily conopeptides.

Many new bioinformatic tools have been created within the Conco project to help with the conopeptide classification and characterisation. Many of these tools were part of the PhD thesis of Dominique Koua (see chapter 6.2).

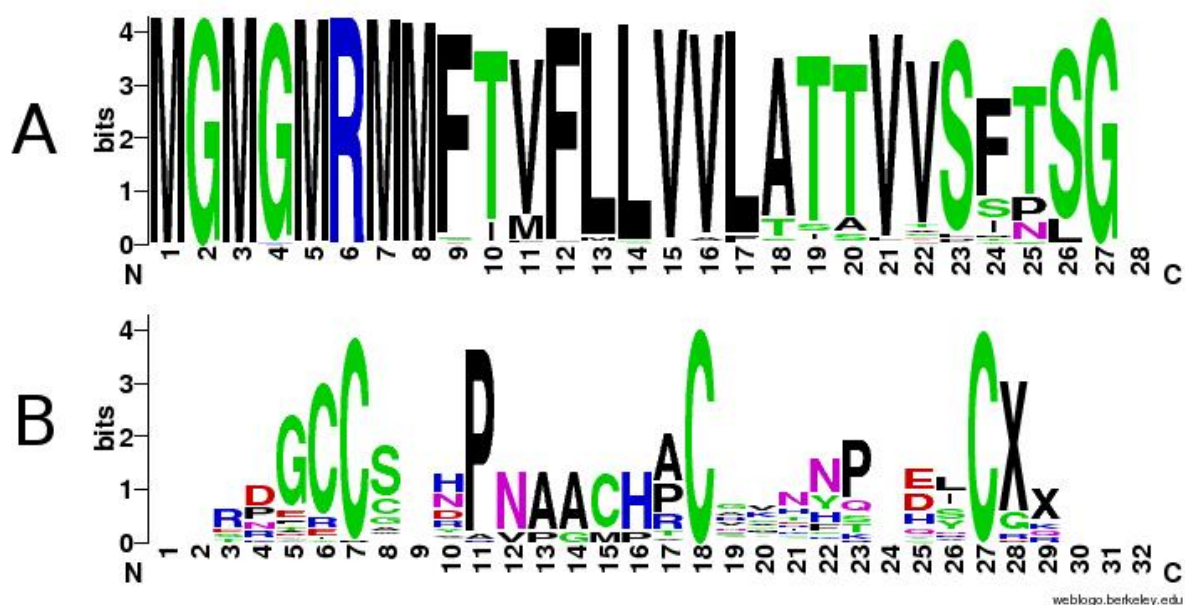


Figure 13: Amino acid conservation level comparison.

A) High amino acid conservation level within the signal sequences of the gene superfamily A. **B)** Conversely, mature sequence conopeptides show high variability in nearly all amino acid positions, except mainly for the cysteine residues. These graphics are based on the sequence alignment, respectively, of all the superfamily A signal sequences and of all the superfamily A alpha-conopeptide mature sequences deposited to date into the Conoserver database. Source for the logo creation: <http://weblogo.berkeley.edu/>

3.3 Conopeptides nomenclature

Historically, the nomenclature of conopeptides started with the designation of the conotoxin GI from the *Conus geographus*⁶ where the capital letter indicates the species and the roman number the sequence variant number. Given the growing number of described conopeptides, a more complex standardised nomenclature system had to be put in place. For disulfide-rich conopeptides, a nomenclature based on three aspects was proposed: the first or first two letters indicate the species, the following roman number the cysteine framework and finally an upper case Latin letter points to the order of variants⁷. In some cases, the pharmacological family can be used with a Greek letter at the beginning of the designation (for example α -CnIA) and the synthetic or PTM variants can be designated by IUPAC rules

(for example [Asp1]-CnIH). For disulfide-poor conopeptides (since no cysteine framework can be applied) and conoproteins, the class of peptide or protein (grouped by sequence homology or by similar activity) is used for the nomenclature followed by one or two letters to indicate the species and a number to indicate the order of variants (for example Conomarphin-Cn1).

A specific nomenclature is also used for genetic or transcript sequences. One or two letters are used to identify the species followed by two Arabic numbers separated by a dot, the first for the cysteine framework, the second for the order of variants.

3.4 Conopeptides PTMs

A major feature of conopeptides is the diversity of post-translational modifications (PTMs) they boast. PTMs are known to perform several roles in conopeptides. They are extremely important for the activity of conopeptides and can in many cases, and often through stabilisation, improve their functional efficiency and sensitivity⁸. The most common PTMs found in conopeptides are summarised in Table 2. The importance of PTMs for the activity of conopeptides has been extensively demonstrated through many studies. However, the exact role of PTMs in the interaction mechanisms of the conopeptides with their privileged target still needs to be clarified. Due to the many possible PTM combinations, PTMs can also be used to characterize different subunit variations in some conopeptide receptors^{9,10}.

Beyond the sole case of conopeptides, the characterisation of post-translationally modified peptides in general is quite challenging. Even though mass spectrometry techniques have noticeably improved over the past decades, the correct detection and characterisation of some PTMs are still problematic. The instruments themselves are not really at fault but rather more the processing of their data. While the development of new and more efficient

3. Conopeptides and Conoproteins

bioinformatic tools have been of great help in this respect, many more improvements are possible and needed to enhance the proteomic results¹¹.

Table 2: List of the most common post-translational modifications observed in conopeptides^{9,10}

Modification name	Mass shift (Da)	Modification site(s)	Function(s)
Amidation	-1	C-terminus following removal of Gly-Lys or Gly-Arg	Affects binding affinity
Bromination	+78	Trp	Affects bioactivity
Carboxylation	+44	Glu	Stabilises α -helix formation and is a Ca^{2+} ligand
Disulfide bond formation	-2	Cys	Confers stability
L- to D-epimerization	0	Trp, Phe, Val	Confers stability and resistance to enzymatic degradation
Glycosylation	Varies, can be a combination of sugar units like hexose (+162) and/or N-acetyl-hexosamine (+203)	Ser, Thr	Confers stability and facilitates transport to target
Hydroxylation	+16	Pro, Val	Confers stability and enhances binding through hydrogen bonds
Pyroglutamylation	-17 (-18)	Gln (Glu)	Confers stability through N-terminal blocking
Sulfation	+80	Tyr	Confers stability and solubility

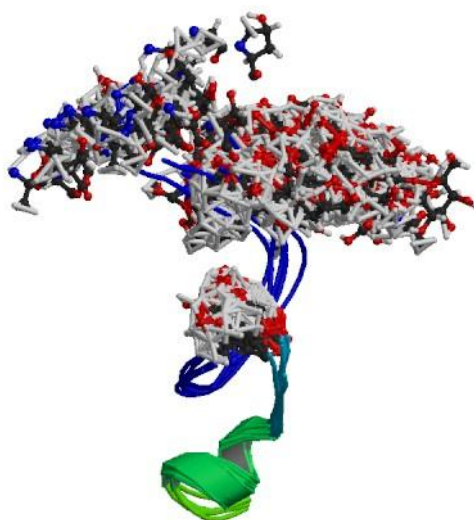


Figure 14: NMR structure of the CcTx conopeptide with O-glycosylation PTM.

Although the complex structure of glycans clearly outstands, its exact role in the conopeptide activity is not known, but it could be responsible for enhancing its stability¹².

3.5 Conopeptides targets

The diversity, as well as the pharmacological and therapeutic potentials of conopeptides are immense. The structure-activity relationship studies have demonstrated the variable efficiency, potency and sensitivity of conopeptides on many different targets^{3,4,13,14}. Not only have conopeptides become interesting for new potential drug candidates but they have proven their excellent aptitudes as pharmacological tools for the discovery and study of many different pharmacological targets. In the case of neurological targets, 12 pharmacological families have been described to date (see Table 3).

3. Conopeptides and Conoproteins

Table 3: Summary of the 12 conopeptide pharmacological families described to date.

Conopeptide family	Type of target	Brief mode of action	Possible applications	Ref.
α (alpha)	nicotinic acetylcholine receptors (nAChR)	Specifically inhibits subtypes of nAChRs	Mainly pain, but also neuronal disorder diseases such as epilepsy, schizophrenia, nicotine addiction, Alzheimer's and Parkinson's disease	15
γ (gamma)	Neuronal pacemaker cation currents	Triggers action potential bursting	Epilepsy and pacemaker potentials in heart muscle	16
δ (delta)	Voltage-gated Na channels (Na_v)	Agonist; inhibits fast Na_v inactivation	Arrhythmias, epilepsy	17
ϵ (epsilon)	Presynaptic Ca^{2+} channels or G protein-coupled pre-synaptic receptors (GPCR)	Reduces the presynaptic influx of Ca^{2+} in a slow and reversible fashion	Presynaptic linked neuronal disorder diseases	18
i (iota)	Voltage-gated Na channels (Na_v)	Agonist; shifts the voltage dependence of activation	Heart failures and pain	19
κ (kappa)	Voltage-gated K channels (K_v)	Blocker; inhibits the Shaker K_v channel	Neuronal disorder diseases and cancer	20
μ (mu)	Voltage-gated Na channels (Na_v)	Antagonist, blocker; reversibly inhibits Na_v channel	Pain, stroke, epilepsy	21
ρ (rho)	$\alpha 1$ -adrenoceptors (GPCR)	non-competitively inhibits $\alpha 1$ -adrenoceptors	Cardiovascular diseases, benign prostatic hypertrophy	22
σ (sigma)	Serotonin-gated ion channels (GPCR)	Antagonist; inhibits excitatory serotonin-gated ion channel	Nausea and vomiting; irritable bowel syndrome	23
τ (tau)	Somatostatin receptor (GPCR)	Antagonist	Cancers	24
χ (chi)	Neuronal noradrenaline transporter	non-competitively inhibits norepinephrine uptake	Pain	22
ω (omega)	Voltage-gated Ca channels (Ca_v)	Inhibits Ca_v	Pain, hypertension, arrhythmias, epilepsy	25

3.6 Reference List

1. Violette, A.; Leonardi, A.; Piquemal, D.; Terrat, Y.; Biass, D.; Dutertre, S.; Noguier, F.; Ducancel, F.; Stocklin, R.; Krizaj, I.; Favreau, P. Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar. Drugs* 2012, *10* (2), 258-280.
2. Safavi-Hemami, H.; Gajewiak, J.; Karanth, S.; Robinson, S. D.; Ueberheide, B.; Douglass, A. D.; Schlegel, A.; Imperial, J. S.; Watkins, M.; Bandyopadhyay, P. K.; Yandell, M.; Li, Q.; Purcell, A. W.; Norton, R. S.; Ellgaard, L.; Olivera, B. M. Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl. Acad. Sci. U. S. A* 2015, *112* (6), 1743-1748.
3. Terlau, H.; Olivera, B. M. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol Rev.* 2004, *84* (1), 41-68.
4. Lewis, R. J.; Dutertre, S.; Vetter, I.; Christie, M. J. Conus venom Peptide pharmacology. *Pharmacol. Rev.* 2012, *64* (2), 259-298.
5. Kaas, Q.; Westermann, J. C.; Halai, R.; Wang, C. K.; Craik, D. J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics.* 2008, *24* (3), 445-446.
6. Cruz, L. J.; Gray, W. R.; Olivera, B. M. Purification and properties of a myotoxin from *Conus geographus* venom. *Arch. Biochem. Biophys.* 1978, *190* (2), 539-548.
7. Gray, W. R.; Olivera, B. M.; Cruz, L. J. Peptide toxins from venomous *Conus* snails. *Annu. Rev. Biochem.* 1988, *57* (-), 665-700.
8. Buczek, O.; Bulaj, G.; Olivera, B. M. Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol. Life Sci.* 2005, *62* (24), 3067-3079.
9. Jakubowski, J. A.; Kelley, W. P.; Sweedler, J. V. Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon* 2006, *47* (6), 688-699.
10. Craig, A. G.; Bandyopadhyay, P.; Olivera, B. M. Post-translationally modified neuropeptides from *Conus* venoms. *Eur. J. Biochem.* 1999, *264* (2), 271-275.
11. Ahrne, E.; Muller, M.; Lisacek, F. Unrestricted identification of modified proteins using MS/MS. *Proteomics.* 2010, *10* (4), 671-686.

12. Hocking, H. G.; Gerwig, G. J.; Dutertre, S.; Violette, A.; Favreau, P.; Stocklin, R.; Kamerling, J. P.; Boelens, R. Structure of the O-Glycosylated Conopeptide CcTx from *Conus consors* Venom. *Chemistry*. 2013, 19 (3), 870-879.
13. Lewis, R. J.; Garcia, M. L. Therapeutic potential of venom peptides. *Nat. Rev. Drug Discov.* 2003, 2 (10), 790-802.
14. Vetter, I.; Dekan, Z.; Knapp, O.; Adams, D. J.; Alewood, P. F.; Lewis, R. J. Isolation, characterization and total regioselective synthesis of the novel muO-conotoxin MfVIA from *Conus magnificus* that targets voltage-gated sodium channels. *Biochem. Pharmacol.* 2012, 84 (4), 540-548.
15. Lebbe, E. K.; Peigneur, S.; Wijesekara, I.; Tytgat, J. Conotoxins targeting nicotinic acetylcholine receptors: an overview. *Mar. Drugs* 2014, 12 (5), 2970-3004.
16. Fainzilber, M.; Nakamura, T.; Lodder, J. C.; Zlotkin, E.; Kits, K. S.; Burlingame, A. L. gamma-Conotoxin-PnVIIA, a gamma-carboxyglutamate-containing peptide agonist of neuronal pacemaker cation currents. *Biochemistry* 1998, 37 (6), 1470-1477.
17. Leipold, E.; Hansel, A.; Olivera, B. M.; Terlau, H.; Heinemann, S. H. Molecular interaction of delta-conotoxins with voltage-gated sodium channels. *FEBS Lett.* 2005, 579 (18), 3881-3884.
18. Rigby, A. C.; Lucas-Meunier, E.; Kalume, D. E.; Czerwiec, E.; Hambe, B.; Dahlqvist, I.; Fossier, P.; Baux, G.; Roepstorff, P.; Baleja, J. D.; Furie, B. C.; Furie, B.; Stenflo, J. A conotoxin from *Conus textile* with unusual posttranslational modifications reduces presynaptic Ca²⁺ influx. *Proc. Natl. Acad. Sci. U. S. A* 1999, 96 (10), 5758-5763.
19. Buczek, O.; Wei, D.; Babon, J. J.; Yang, X.; Fiedler, B.; Chen, P.; Yoshikami, D.; Olivera, B. M.; Bulaj, G.; Norton, R. S. Structure and sodium channel activity of an excitatory I1-superfamily conotoxin. *Biochemistry* 2007, 46 (35), 9929-9940.
20. Terlau, H.; Shon, K. J.; Grilley, M.; Stocker, M.; Stuhmer, W.; Olivera, B. M. Strategy for rapid immobilization of prey by a fish-hunting marine snail. *Nature* 1996, 381 (6578), 148-151.
21. Cruz, L. J.; Gray, W. R.; Olivera, B. M.; Zeikus, R. D.; Kerr, L.; Yoshikami, D.; Moczydlowski, E. *Conus geographus* toxins that discriminate between neuronal and muscle sodium channels. *J. Biol. Chem.* 1985, 260 (16), 9280-9288.
22. Sharpe, I. A.; Gehrmann, J.; Loughnan, M. L.; Thomas, L.; Adams, D. A.; Atkins, A.; Palant, E.; Craik, D. J.; Adams, D. J.; Alewood, P. F.; Lewis, R. J. Two new classes of

conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter. *Nat. Neurosci.* 2001, 4 (9), 902-907.

23. England, L. J.; Imperial, J.; Jacobsen, R.; Craig, A. G.; Gulyas, J.; Akhtar, M.; Rivier, J.; Julius, D.; Olivera, B. M. Inactivation of a serotonin-gated ion channel by a polypeptide toxin from marine snails. *Science* 1998, 281 (5376), 575-578.

24. Petrel, C.; Hocking, H. G.; Reynaud, M.; Upert, G.; Favreau, P.; Biass, D.; Paolini-Bertrand, M.; Peigneur, S.; Tytgat, J.; Gilles, N.; Hartley, O.; Boelens, R.; Stocklin, R.; Servent, D. Identification, structural and pharmacological characterization of tau-CnVA, a conopeptide that selectively interacts with somatostatin sst3 receptor. *Biochem. Pharmacol.* 2013, 85 (11), 1663-1671.

25. Kerr, L. M.; Yoshikami, D. A venom peptide with a novel presynaptic blocking action. *Nature* 1984, 308 (5956), 282-284.

4. Cone snail venom content studies

4.1 Introduction: From the first case of poisoning to the study of conotoxins

As previously mentioned, the venom is produced and matured in the venom duct. The organ itself and its content have been extensively studied due to the remarkable properties of its venom components, the conotoxins. The first report of a case of poisoning by a *Conus* dates back to 1848 by Arthur Adams in the “Narrative of the voyage of the H.M.S. Samarang”¹. He reported that Sir Edward Belcher, Captain and Commander of the expedition, “was bitten by one of these Cones, which suddenly exerted its proboscis as he took it out of the water with his hand”. The observation continued by describing the tool used by the cone: “The instrument which inflicted the wound, in this instance, I conceive must have been the tongue, which in these mollusks, is long, and armed with two ranges of sharp-pointed teeth”. But at the time the venom had not been studied yet. Other cases of injuries or deaths due to marine snails of the genus *Conus* were reported since²⁻⁵ and thus cone snails became an intriguing subject of study not only for the beauty of their shell, but also for the potency of their venom. One of the first biochemistry studies of cone snail venom dates back to 1960 when Kohn provided the first clues that the cone snail venom is a mixture of several active components⁶. At the beginning of the ‘60s, several studies on the pharmacology of the cone snail venom followed⁷⁻⁹ and it was thought that the venom was maybe composed of large molecules or of proteins. The first study proving that proteins are responsible for the venom’s toxicity was published in 1975 by Olivera’s group¹⁰. Their work provided evidence of a higher protein content than expected since it was estimated between 5% and 25%. Of course, at that time, techniques for protein quantification did not, or not correctly, detect peptide content, meaning

that the large portion of peptide content, now known to be responsible for the venom toxicity, was still not known. On this last point though, Cruz *et al.* strongly suspected that smaller and several components per venom could be responsible for the toxicity, and that these compounds seemed to be quite resistant to heat and to enzyme digestion. The next logical step was purification.

From this point onwards - in parallel to the study of the pharmacological effects of cone snail venom, undertaken mostly in Australia ¹¹⁻¹⁶ - the interest in both its composition and the extension of the studies to other species started growing fast and worldwide. One of the first biochemical studies of the venom was undertaken by Olivera's group where they managed to purify a 13 amino acid, cysteine rich, peptide responsible for the toxic activity ¹⁷. At this stage, it clearly appeared that cone snail venoms could be a valuable source of bioactive components, particularly bearing in mind that more than 300 species of this genus were described at that time. This work also paved the way to the cone snail venom toxinology nomenclature still in use today, where the term of "conotoxin" was suggested for the first time when speaking of those toxic peptides. They also used the first letter of the species followed by a roman number to classify potential series of peptides in the same species. For instance, the new peptide was called "Conotoxin GI", the first peptide described from a *Conus geographus*.

4.2 The early stages of peptidomics

Since these first biochemical studies, the interest in the study of cone snail venoms grew exponentially. In parallel to pharmacological studies of few purified conotoxins ¹⁸⁻²¹, a few studies and reviews on cone snail venom content in general appeared in the early '80s, trying to understand the global mechanism of the venomous function of cone snails and the pharmacological potential of their venom ^{22,23}. At that time, it was thought that cone snail

venoms contained only a few dozens components. It was only in the mid '90s that the use of new techniques in mass spectrometry for the study of cone snail venom appeared ²⁴⁻²⁶. Mass spectrometry became a *sine qua non* for the characterization of new components and the proper elucidation of the primary structure, including several post-translational modifications (PTMs). At this point, these studies revealed that the possible PTMs were clearly underestimated on conotoxins, as well as the number of components per venom, which were then estimated at about a hundred.

While the trend and future insights on venom peptidomics appeared in the 2000s ²⁷, it was only in the mid 2000s that the first peptidomic studies using advanced mass spectrometry techniques fully revealed the complexity of cone snail venom content. Indeed, the high number of possible PTM was confirmed and new insights on the venom sampling and analysis were provided. Several studies began to insist on the importance of different mass spectrometry techniques and different venom sources to reveal the full potential of the different cone snail venoms ²⁸⁻³². From that moment hence, mass spectrometry became a *sine qua non* for the study of the full peptide and protein content of all kinds of venoms. This paved the way to venom peptidomics or venomics ³³.

4.3 Peptidomics of a fish hunting cone snail

Within the framework of this “venomics” project ³³, an extensive study of the venom content of a fish-hunting cone snail, *Conus consors*, was launched. Using the latest chromatography and mass spectrometry techniques, the aim was to improve the understanding of the function and evolution of the cone snail venomous system, shed more light on the diversity and evolution of venom peptides and proteins, and possibly contribute to new developments in the field of drug discovery.

In the following article, an optimisation of the liquid chromatography separation step is discussed as well as the comparison of two different mass spectrometry techniques - MALDI-MS and ESI-MS. The pros and cons of these techniques applied to the study of cone snail venoms are also briefly discussed.

Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*

Daniel Biass¹, Sébastien Dutertre¹, Alain Gerbault², Jean-Louis Menou³, Robin Offord⁴,
Philippe Favreau¹ and Reto Stöcklin^{1&2}

¹Atheris Laboratories, case postale 314, CH-1233 Bernex-Geneva, Switzerland;

²Toxinomics Foundation, Nouméa branch, chemin des Aulx 18, CH-1228 Plan-les-Ouates, Geneva, Switzerland;

³Institut de Recherche pour le Développement, BPA5, 98848 Nouméa cedex, New Caledonia;

⁴Mintaka Foundation for Medical Research, chemin des Aulx 14, CH-1228 Plan-les-Ouates, Geneva, Switzerland.

Publication data

Published the 06.03.2009 in the Journal of Proteomics, Vol. 72, Issue 2, p.210-218.

Pubmed ID: 19457347

DOI: 10.1016/j.jprot.2009.01.019

Cited 49 times on Google Scholar, 41 times on ScienceDirect.

Keywords

cone snail, conopeptide, conotoxin, *Conus consors*, ESI-MS, HPLC, hydrophobicity, MALDI-TOF-MS, mass fingerprint, mass range, mass spectrometry, peptide, peptidomics, proteomics, toxins, venom, venomics.

Personal contribution

All the work from sampling to data analysis and the manuscript writing.

Abstract

In the context of an exhaustive study of the piscivorous cone snail *Conus consors*, we performed an in-depth analysis of the intact molecular masses that can be detected in the animal's venom, using MALDI and ESI mass spectrometry. We clearly demonstrated that, for the venom of this species at least, it is essential to use both techniques in order to obtain the broadest data set of molecular masses. Only 20% of the total number of molecules detected were found in both mass lists. The two data sets were also compared in terms of mass range and relative hydrophobicity of the components detected in each. With a view to an extensive analysis of this venom's proteome, we further performed a comparative study by ESI-MS between venom obtained after classical dissection of the venom duct *versus* venom obtained by milking live animals. Surprisingly, although many fewer components were found in the milked venom than in the dissected venom, ~50% of those found had not been seen in the dissected venom. Several questions raised by these observations are discussed. With regards to the current knowledge of the cone snail venom composition, our results emphasize the complementary nature of the mass spectrometry methods and of the two techniques used in venom collection.

1. Introduction

Animal venoms are highly complex mixtures of biologically active compounds. These biomolecules have been tailor-made by millions of years of evolution to endow their possessors with the means to carry out the specific offensive and defensive tasks needed for their survival (Olivera, 1997). Venom molecules were thus gradually modified to match a multitude of highly specific targets, hence their unique pharmacological properties. Thanks to the resemblance of these targets to mammalian receptors, a remarkable number of venom components has been successfully developed as new research tools and therapeutic drugs (Lewis & Garcia, 2003). Unfortunately, despite the large number of venomous animals and the complexity of their venoms, only a tiny proportion (estimated to represent less than 0.1%) of venom components have been identified and characterized, and less than 1% of genetic information is available (Menez, Stocklin, & Mebs, 2006). While this low percentage may reflect the sparseness of some venomous animals, it also stems from the bioactivity-guided research approaches traditionally implemented to find new bioactive molecules. Indeed, very often, a biological activity assay is the first step taken in the quest for new compounds, followed by the isolation and characterization of the native bioactive substance from natural libraries. This strategy is time consuming and requires large amounts of material. Nowadays, current mass spectrometry (MS) techniques can generate an abundance of valuable data not only in a very short period of time, but more importantly using much smaller sample amounts (Favreau et al., 2006). Through a structure-driven process and thanks to constant evolving biocomputing capacities, MS has become paramount not only for analytical purposes, but also for the rapid discovery and characterization of new components in the field of toxinology (Stocklin & Favreau, 2002). When used in combination with DNA sequencing from cDNA libraries or ESTs, a wealth of information on the venom gland components can be obtained (Sanz et al., 2008; Escoubas, Sollod, & King, 2006; Gowd, Dewan, Iengar, Krishnan, & Balaram, 2008).

Different types of mass spectrometers are used to untangle the complexity of venom mixtures and to rapidly produce a large amount of information, such as the molecular masses of intact components, the number of disulfide bridges, and primary sequences. Not only do the accuracy and sensitivity of these instruments now enable us to distinguish between species on the basis of venom composition (for quality control of individual venom batches or for taxonomic or phylogenetic studies), but they can also reveal intersexual and other intraspecific variations (Creer et al., 2003; Calvete, Escolano, & Sanz, 2007; Herzig & Hodgson, 2008). Although MALDI-MS equipped with a time-of-flight (TOF) mass analyzer proves to be particularly well suited to study complex venom compositions, this technique suffers from a low dynamic range, ion suppression effects and poor resolution in the linear mode for high masses (Escoubas, Celerier, & Nakajima, 1997). Furthermore, the introduction of an off-line RP-HPLC step with fraction collection and freeze-drying prior to the analysis by MALDI-TOF-MS and ESI-MS of isolated fractions adds an important increasing factor to mass detection (Pimenta, Stocklin, Favreau, Bougis, & Martin-Eauclaire, 2001; Perchuc, Menin, Stocklin, Buhler, & Schoni, 2005). Most of these studies also tend to prove that, in many cases, the number of components present in venoms is consistently underestimated. The real size and variability of individual venom proteomes remain an open question.

In this study, we propose two comparative analyses of complex venom mixtures that initiate an in depth venom project of a marine venomous organism, namely *Conus consors* (Menez et al., 2006). We first present a comparison between MALDI-MS and ESI-MS analyses of individual RP-HPLC fractions of this fish-hunting (piscivorous) cone snail venom to determine the extent of overlap and/or complementarities between these two techniques in terms of mass detection. We also evaluate to what extent a given technique is better suited for the detection of high *versus* low masses and hydrophilic *versus* hydrophobic components present in these specific venoms. So far, studies on cone snail venom have always relied on dissected material, with very few exceptions. However, a recent study has shown that the injected venom is significantly different to that of dissected preparations (Jakubowski, Kelley,

Sweedler, Gilly, & Schulz, 2005). We therefore also propose a direct comparison of all the masses detected using ESI-MS, between on the one hand, dissected *Conus consors* venom and on the other hand, venom milked from live specimens of the same species.

2. Materials and methods

Acetonitrile (ACN, from Fisher Scientific Ltd., Loughborough, UK), trifluoroacetic acid (TFA, from Pierce-Perbio, Lausanne, Switzerland) and formic acid (Acros Organics, Geel, Belgium) were of HPLC gradient grade or higher. Deionised water was purified using a Milli-Q system (Millipore Corp., Billerica, MA, USA). When needed, each solvent prepared was filtered and sonicated before use.

2.1 Venom preparations

All 25 specimens of *Conus consors* used for this study were collected from one colony in the Chesterfield Islands (New Caledonia) in the frame of the CONFIELD scientific expedition in June 2007. The pool of crude venom, referred to as dissected venom (DV) further in this article, was obtained after dissection of 19 *Conus consors* specimens following the previously described method (Favreau et al., 1999). The crude venom pool has been lyophilized and weighed 35.8 mg. The protein content of this pool was estimated at about 7 mg (20%). For proteomic analyses, the lyophilized dissected venoms pool was reconstituted at 1 mg/mL (protein content) in acidified water (0.1% TFA) and desalted using solid-phase extraction onto a Sep-Pak Vac 35cc cartridge equilibrated in acidified water according to the manufacturer's instructions (Waters, Milford, MA, USA). Elution was performed with 70% ACN in acidified water and the eluate freeze-dried under vacuum in a SpeedVac concentrator (ThermoSavant, Holbrook, NY, USA), then stored at -80°C. The pool of milked venom (MV) was obtained by combining 67 individual milkings from a batch of 6 specimens

kept alive in our aquariums. The milked venom was lyophilized and stored at -80°C. Its total protein content was estimated at about 400 µg.

2.2 RP-HPLC

Reversed-phase high-performance liquid chromatography (RP-HPLC) was carried out using a Waters Alliance 2795 system equipped with a Waters 996 Photodiode Array Detector under control of the Waters Millennium³² 4.0 software (Waters, Milford, MA, USA). All fractionations were performed using a 218TP510 Protein and Peptide C₁₈ RP column (10 mm internal diameter / 250 mm length, from Vydac, Hesperia, CA, USA) with a gradient combining solvent A (0.1% TFA in water) and solvent B (90% ACN / 0.1% TFA in water). A flow rate of 2.0 mL/min was used with a gradient of 1% B per minute, starting from 100% of solvent A. UV detection of the fractions was carried out at 214 nm and fractions were collected manually. The dissected venom pool was subjected to 5 RP-HPLC runs and corresponding fractions were pooled. The milked venom pool was fractionated in a single RP-HPLC run. All final fractions were freeze-dried and stored at -80°C.

2.3 MALDI-MS

MALDI-MS analyses were carried out on an Ultraflex TOF-TOF mass spectrometer operated in positive reflector and linear modes under control of the FlexControl 2.2 software (Bruker, Bremen, Germany). 2% of dissected venom fractions were dissolved in 10 µL of solvent (H₂O/ACN/TFA, 79.92:20:0.08, v/v/v). In reflector mode (480-5000 Da), samples were deposited on a 384 AnchorChip 600 plate (Schuerenberg et al., 2000) using the affinity preparation method (Gobom et al., 2001). In short, the plate was pre-coated with alpha-cyano-4-hydroxycinnamic acid (CHCA, 0.8 mg/mL in acetone/aqueous 0.1% TFA 97:3) and 0.5 µL of the venom sample was deposited at 3 different dilutions 1:10, 1:100 and 1:1000.

Re-crystallization was performed by addition of 0.25 μL of CHCA (0.5 mg/mL in ethanol/acetone/aqueous 0.1% TFA 60:30:10). External calibration was carried out with a mixture of 9 peptides in the 700–3500 Da mass range with a maximum deviation tolerance of 50 ppm. In linear mode (1000–75000 Da), samples were deposited on a 384 polished steel plate using the sandwich method: the plate was pre-coated with CHCA saturated in acetone and 0.5 μL of the venom sample was deposited at 3 different dilutions 1:10, 1:100 and 1:1000. Re-crystallization was performed by addition of 0.5 μL of CHCA (saturated in acetone/aqueous 0.1% TFA (50/50)). External calibration was carried out with a mixture of 4 proteins in the 5000–20000 Da mass range with a 250 ppm maximum deviation tolerance. The FlexAnalysis software was used for data processing and to assist data analysis.

2.4 ESI-MS

ESI-MS analyses were performed on a Q-TOF micro mass spectrometer (Waters-Micromass, Manchester, UK) equipped with its standard ESI source and operated under control of the MassLynx 4.0 software (Waters-Micromass). A portion of 3% of each dissected venom fraction was dissolved in 20 μL of solvent ($\text{H}_2\text{O}/\text{ACN}/\text{HCOOH}$ 49.8:50:0.2, v/v/v). A portion of 20% of each milked venom fractions was dissolved in 300 μL of the same solvent. Acquisitions were carried out from 100 to 1800 m/z in 1 sec. in the positive ionization mode by sample loop injection of 10 μL of the analyte with infusion at a flow rate of 5 $\mu\text{L}/\text{min}$. The MassLynx software was used for data processing and to assist data analysis. External calibration of the instrument was performed with glu-fibrinopeptide-B with a maximum deviation tolerance of 50 ppm.

2.5 Bioinformatics

All mass list comparisons were carried out automatically using an in-house programme allowing the generation of a unique list of all detected molecular masses that overlap within

50 ppm (for all ESI-MS and MALDI-TOF-MS in reflector mode) or 250 ppm (for MALDI-MS in linear mode).

3. Results and discussion

The initial aim of this study was to compare, in the most exhaustive way, mass detection capabilities and properties between two soft ionization mass spectrometry methods, MALDI-MS and ESI-MS, on a complex mixture of compounds present in the venom of the piscivorous cone snail *Conus consors*.

3.1 RP-HPLC

Prior to MS analysis of the *Conus consors* dissected venom, a reversed-phase chromatography was carried out to fractionate the venom components. This first separation step provided good peak resolution (and good reproducibility, facilitating the fraction pooling between runs). The profile of the chromatogram reveals a relatively complex venom, as described in the case of other piscivorous cone snail venoms, namely, *C. geographus* (Olivera et al., 1985), *C. striatus* (Ramilo et al., 1992) or *C. achatinus* (Gowd et al., 2008). It highlights the presence of three major groups of components according to their hydrophobicity (~15 to 45 min., 50 to 70 min. and 75 to 85 min. of retention time). For the dissected venom, a total of 108 fractions were collected manually.

Since the initial aim of this study was to detect a maximum number of masses from this cone snail venom, we optimized our protocol to achieve the highest possible sensitivity during MS analyses. Often in ESI-MS, RP-HPLC is directly coupled to the MS, which is faster and more convenient. However, depending on the mobile phase and analytical conditions, different sensitivities have been observed during MS analyses (Garcia, 2005). Unfortunately, it appears that no satisfactory elution solvent can afford high LC peak resolution with optimal ESI-MS sensitivity. An optimized RP-HPLC step allows fraction storage and availability for

further analyses. Fractions can be lyophilized and dissolved not only in the most appropriate solvent for subsequent analysis, but also at a higher concentration than in the original fraction. This can strongly increase the performance of an MS analysis, which is particularly true in the case of ESI-MS. For all these reasons, a liquid chromatography step was performed off-line prior to ESI-MS and MALDI-TOF-MS analyses of the LC fractions. Importantly, this strategy enables an effective comparison between the two MS techniques considered here, since the samples undergo the same pre-analytical process and therefore the same material is used as the starting point.

3.2 MS analyses

As the quality of MALDI-TOF-MS results largely depends on the sample preparation technique (Padliya & Wood, 2008), and since venoms are complex mixtures that are highly variable among species, it was essential to first evaluate the most adequate sample preparation method. To this end, preliminary analyses were undertaken with various matrices, deposition methods and plates using cone snail venoms (data not shown). In our case, the affinity preparation with alpha-cyano-4-hydroxy-cinnamic acid on AnchorChip plates led to the best results in reflector mode (m/z 480-5000 Da). This combination appeared to be the most adequate for the analysis of crude venom fractions, resulting in high mass accuracy, high resolution, good signal-to-noise ratio and decreased ion suppression effect. In linear mode (m/z 1000-75000 Da), several sample preparation techniques were also tested, resulting in the selection of the classical sandwich deposition method on polished steel plate for best results.

For ESI-MS analyses, instrumental parameters were optimized for high sensitivity with the use of a standard ESI solvent containing 0.2% formic acid. Since formic acid is a weak acid, the ion-pairing is known to be much better suited for the electrospray ionisation conditions than with TFA, which induces a much stronger ion-pairing (Apffel, Fischer, Goldberg,

Goodley, & Kuhlmann, 1995). Formic acid, which is a much less efficient counter-ion for RP-HPLC, enables a much better signal-to-noise ratio than TFA.

3.3 MALDI-MS *versus* ESI-MS

Analysis of the venom fractions revealed a total of 889 components using MALDI-MS, while 1078 components were detected by ESI-MS. It is important to note that manual processing of all measured molecular masses was performed in order to remove all putative Na⁺, K⁺ and Fe²⁺ adducts from the final mass lists. Furthermore, these figures do not include masses below 800 Da to stay within conditions that allow a rigorous comparison and avoid possible attributions of matrix or solvent signals by MALDI-MS. ESI-MS and MALDI-MS in reflector mode enabled good molecular mass measurement accuracy ($\pm$ 50 ppm), whereas resolution and accuracy were poorer with MALDI-MS in linear mode. Masses common to both techniques represent only 20.8% (339 components) of the total number of compounds detected (Figure 15, inset). For these masses, there was no direct correlation with the mass intensities recorded from the MALDI-MS. In other words, the common masses are not necessarily the most abundant signals generated by the venom components.

The molecular mass range is asymmetrically distributed, with about 60% of the components showing a mass comprised between 1000 to 3000 Da. In the 1000-2000 mass range, masses detected with both techniques represent about half of the masses detected per single technique, exponentially decreasing at higher mass ranges (Figure 15). In addition, within this 1000-2000 mass range, MALDI-MS tends to detect more components than ESI-MS, whereas the contrary occurs with masses above 5000 Da, although the differences are not significant (Figure 16). As for hydrophobicity, components appear to be evenly detected by ESI-MS throughout the full range, whereas MALDI-MS shows a better detection in the lower hydrophobicity range (Figure 16, inset).

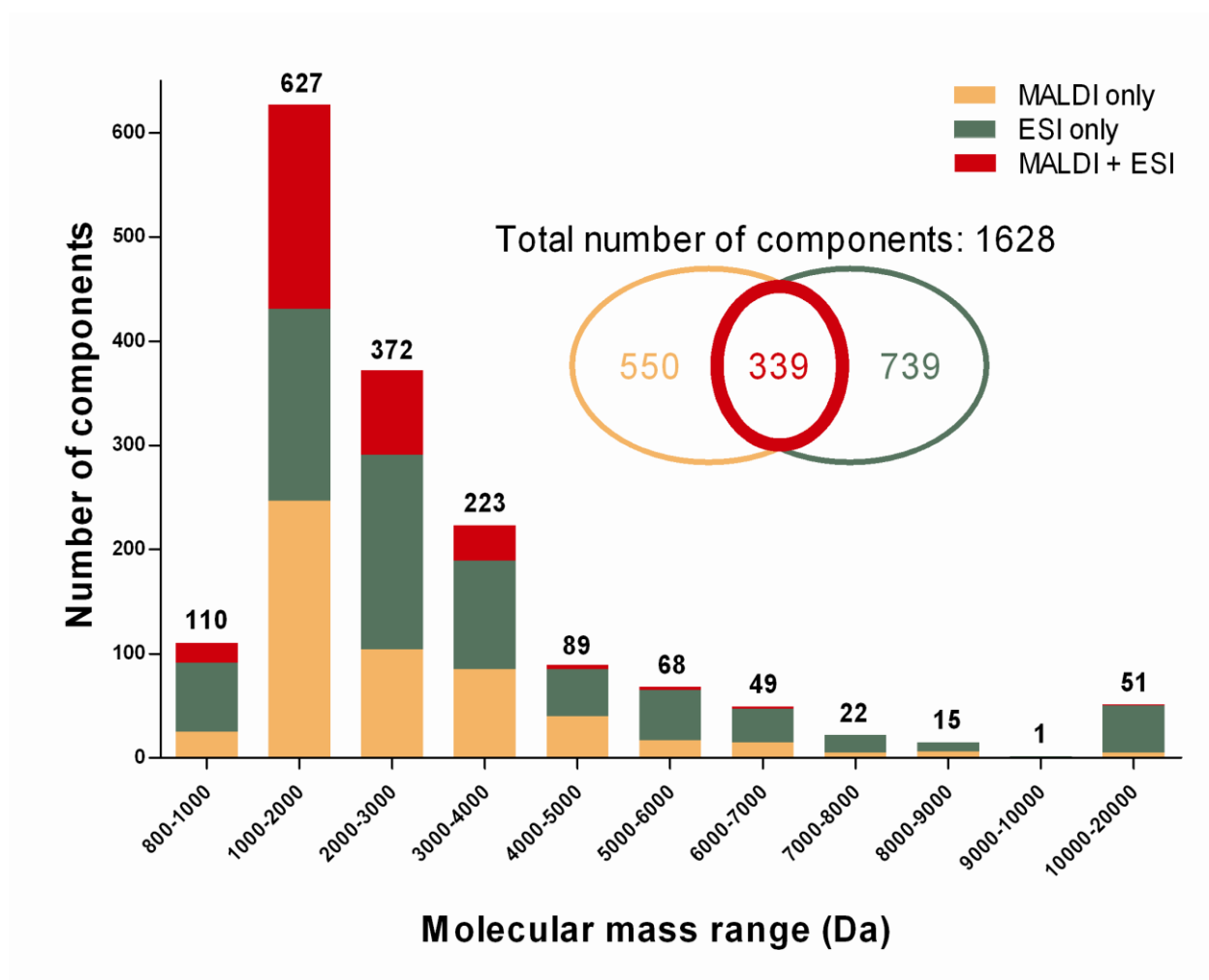


Figure 15: Distribution of *Conus consors* dissected venom fractions molecular masses obtained by ESI-MS and MALDI-MS.

Overlaid histograms represent the frequency of molecular mass distribution per technique used. For a total of 1628 measured masses, 550 compounds are detected only by MALDI-MS, 739 only by ESI-MS and 339 are common to both techniques, demonstrating that the overlap between the two techniques represents only about 20% of all the masses. A left asymmetrical overall distribution is visible with about 60% of the components in the 1000 to 3000 mass range. Masses in common represent about half of the masses detected per single technique in the 1000-2000 mass range, exponentially reducing in higher mass ranges.

Even if MALDI-MS is more sensitive than ESI-MS in terms of detection level, the current results demonstrate that slightly more masses were detected using ESI-MS than with MALDI-MS. A recent study on the Indian red scorpion *Mesobuthus tamulus* venom shows

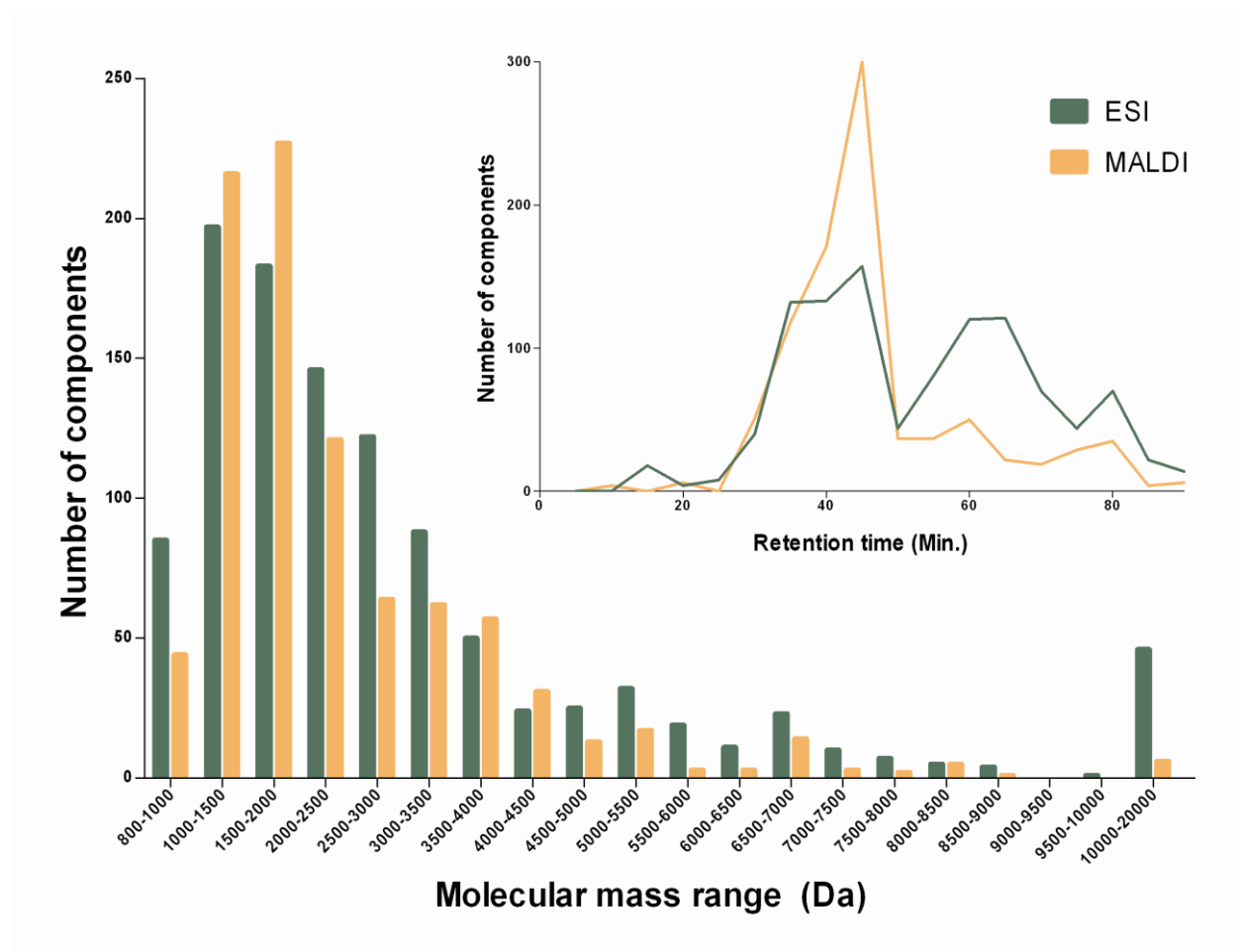


Figure 16: Comparison of *Conus consors* dissected venom fractions between MALDI-MS and ESI-MS techniques according to mass range and hydrophobicity.

The main graphic shows the distribution of the number of compounds with 500 Da mass range windows (except for the first and last classes). One can observe a left asymmetrical distribution with no significant differences between the two techniques. A slight advantage is given to MALDI-MS in the 1000-2000 mass range, whereas ESI-MS seems to have an edge for higher mass ranges as of 4500 Da. The inset graphic represents the frequency of molecular masses found by the two techniques according to the apparent hydrophobicity (retention time) of the compounds. The components appear to be evenly detected by ESI-MS in terms of hydrophobicity, whereas MALDI-MS tends to be better suited in a lower hydrophobicity range.

that, after preliminary off-line size-exclusion chromatography, a larger number of masses are detected by on-line LC/ESI-MS compared to MALDI-MS (Newton, Clench, Deshmukh, Jeyaseelan, & Strong, 2007). Surprisingly, another recent study on the Brazilian scorpion

Opisthacanthus cayaporum venom, describes a better mass detection with MALDI-MS than with on-line LC/ESI-MS after a preliminary off-line liquid chromatography (Schwartz et al., 2008). Mass detection therefore can vary, not only according to the components, but also according to the techniques and preparation methods used during the pre-analytical processing, the MS instrument used and the conditions under which it is operated. However, these studies also demonstrate a low mass detection overlap (respectively 37% and 19%, compared to 20% in our study) between MALDI-MS and ESI-MS. This shows to what extent both techniques are complementary and necessary to establish the broadest possible proteome mapping. It is also interesting to note that the asymmetrical mass range distribution found in *Conus consors* venom seems specific to this particular venom. The study on the Brazilian scorpion presents, for example, a trimodal mass range distribution (Schwartz et al., 2008).

3.4 Dissected venom *versus* milked venom

Since we just demonstrated that from an identical starting material, both ESI-MS and MALDI-MS can produce different and complementary results, it appears that two samples can only be reliably compared using one of these methods. We therefore decided to use ESI-MS to conduct a comparison study of dissected venom and milked venom of *Conus consors*. Indeed, direct analysis of dissected and milked venom from the same *Conus* species has rarely been carried out, but these studies have highlighted significant differences (Bingham, Jones, Lewis, Andrew, & Alewood, 1996; Jakubowski et al., 2005). The HPLC conditions for the analysis of the milked venom were the same as those used previously for the dissected venom. A total of 48 fractions were collected manually. Each fraction was analyzed using ESI-MS under the same conditions as for the dissected venom.

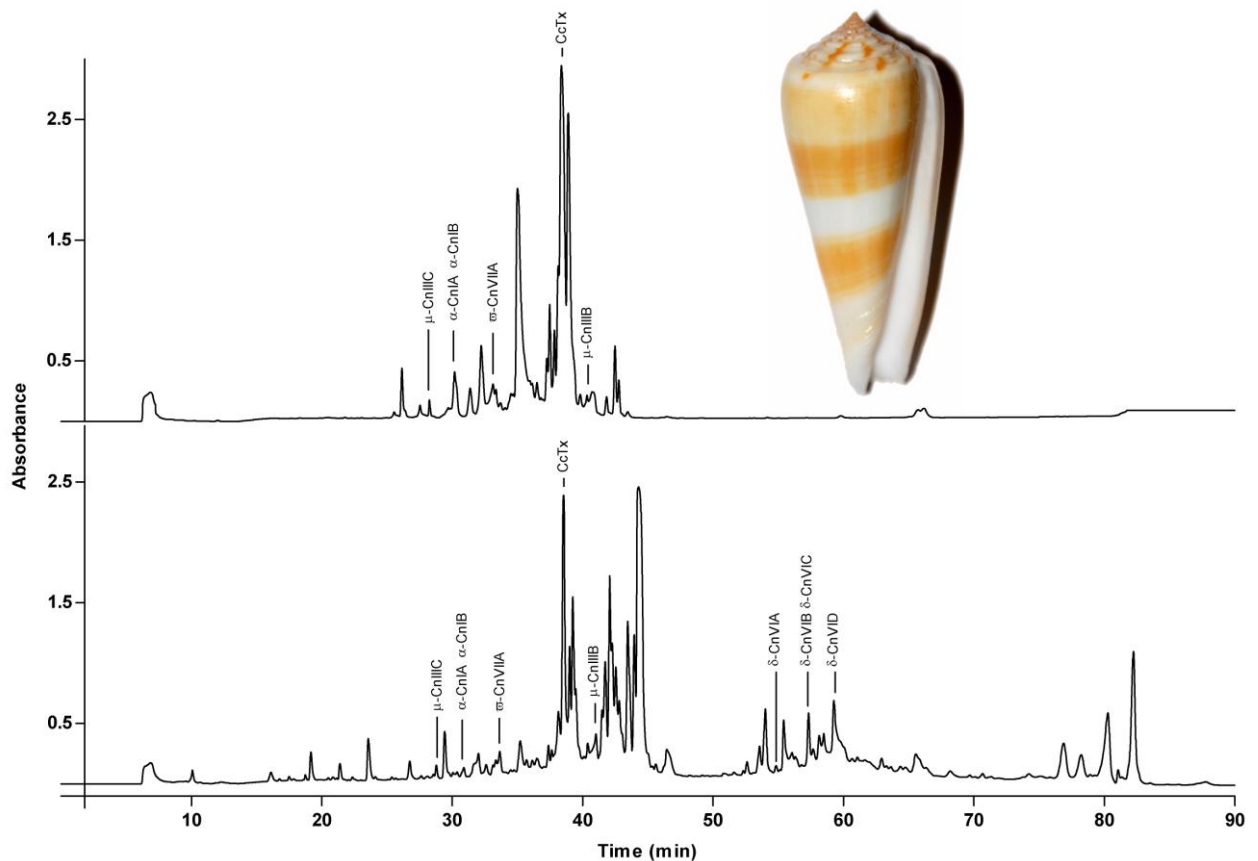


Figure 17: UV chromatograms at 214 nm obtained by RP-HPLC fractionation of milked venom (top) and dissected venom (bottom) of the cone snail, *Conus consors*.

Dissected venom appears to be richer in hydrophobic components than milked venom. Masses matching known *Conus consors* amino acid sequences are pointed out on each chromatogram. Inset image represents a typical shell of *Conus consors* from the Chesterfield Islands.

The chromatograms of the milked and dissected venoms of *Conus consors* show very distinct profiles (Figure 17). Indeed, in the milked venom, much less components are present compared to the dissected venom, and this is particularly evident in the most hydrophobic part of the chromatogram. Compared to other milked cone snail venom profiles, *Conus consors* appears most similar to *Conus striatus* (Jakubowski et al., 2005), but also resembles *Conus purpurascens* (Hopkins et al., 1995) and *Conus ermineus* (Martinez et al., 1995). However, it is much simpler than the milked venom from *Conus obscurus* (Teichert et al., 2004). A complete description of all molecular components detected in the milked venom of

C. consors will be detailed elsewhere (manuscript in preparation). Previously described *Conus consors* peptide sequences were matched to the masses found in both dissected and milked venoms (Table 4). All sequences were found in the dissected venom except for the μ -CnIIIA (Zhang et al., 2006). This conotoxin could not be detected in the milked venom either. A noteworthy fact is that this peptide was described solely on the basis of molecular cloning experiments and never isolated from the venom. Therefore, we cannot exclude that it is not expressed or that post-translational modifications may have occurred in the native peptide, shifting the mass to an unknown value. Since the intensity of both chromatograms is comparable (± 2.5 absorbance at 214 nm), the relative abundance of individual peaks can be compared between milked and dissected venoms. For instance, the previously described conotoxin CcTx appears to be a major peak in both chromatograms (although not the largest in the dissected venom). Its known potent excitatory effect ensures a rapid paralysis of the fish and therefore is essential for prey capture (Le Gall F. et al., 1999). However, the relative amounts of other conotoxins do not seem to correlate. Clearly, it can be seen that a large peak in the milked venom eluting between the ω -conotoxin CnVIIA and the CcTx is almost absent from the dissected venom (Figure 17). Conversely, although largely present in the dissected venom, none of the delta-conotoxins, namely δ -CnVIA, δ -CnVIB, δ -CnVIC and δ -CnVID were found in the milked venom (Bulaj et al., 2001; Favreau, 1999). These conotoxins are indeed part of the major hydrophobic group not present in the milked venom, as can be seen on the RP-HPLC chromatograms (Figure 17). The lack of certain components in the milked venom could result from a specific selection of venom components operated by the animal for venom injection. The underlying mechanism of such a selection is hypothetical, but a study of the anatomy of the *Conus californicus* suggests that a unique type of epithelial cells found at the connection between the venom duct and the pharynx may have selective properties (Marshall et al., 2002). Interestingly, studies on the venom contained in the different portions of a single venom duct show great variations, and the duct portion closest to the pharynx appears to lack the hydrophobic components (Marshall et al., 2002; Garrett,

Buczek, Watkins, Olivera, & Bulaj, 2005). Of course many other factors, such as stress, sex, age or reproductive season (laying eggs) could explain venom composition variations. Additionally, we noticed high intra-specimen variations over a 12 month observation period in the milked venom of *Conus consors* using on-line LC-ESI-MS, with almost no overlap between the conotoxins produced at different times by the same individual (manuscript in preparation). Previous studies on dissected venom between specimens of same location have shown peptide profile variations (Bingham et al., 1996). This phenomenon has also been reported using milked venom (Jakubowski et al., 2005). Even if our analyses did not allow detecting these delta-conotoxins, one cannot exclude their possible presence in very low quantities in the milked venom, which would make them very difficult to detect, even using highly sensitive MS techniques. Indeed, once isolated in minute amounts, these peptides might adhere very easily to the sample tubes, tubings, fittings and columns, thereby possibly getting lost before reaching the ionization source.

Finally, we compared the ESI-MS mass lists of the milked and dissected venoms of *Conus consors*. We found a total of 150 components in the milked venom which is, as mentioned previously, poor compared to the total of 1078 compounds that were detected in the dissected venom. Yet, since most of the conotoxins previously isolated from the dissected venom duct based on bioassays were also detected in the milked venom (Table 4), the reduced complexity does not necessarily mean a less interesting material. Importantly, milking avoids sacrificing the animals, and long term production of milked venom can provide amounts of toxins large enough for complete biochemical and pharmacological characterizations (Hopkins et al., 1995; Kelley, Schulz, Jakubowski, Gilly, & Sweedler, 2006; Schroeder et al., 2008). In addition, a less complex venom facilitates the isolation of the active fraction. Finally, the observed higher complexity of the dissected venom most likely derives from the presence of cellular debris, unprocessed toxin precursors, propeptides, and/or degradation fragments. Therefore, while at first glance the milked venom appears less complex than the dissected venom, and one could fear missing on important molecules, we

propose that the milked venom might contain a judiciously pre-selected set of highly biologically relevant toxins: those that are used to capture specific preys.

Table 4: Previously described *Conus consors* venom sequences, their calculated monoisotopic masses and matching masses measured by ESI-MS in the dissected venom (DV) and milked venom (MV).

Conotoxin	Sequence	Calculated mass (Da)	Measured mass in DV (Da)	Measured mass in MV (Da)	Ref.
α -CnIA	GRCCHPACGKYYS*	1541.57	1541.53	1541.57	(Favreau et al., 1999)
α -CnIB	--CCHPACGKYYS*	1328.45	1328.43	1328.44	(Favreau et al., 1999)
μ -CnIIIA	GRCCDVPNACSGRWCRDHAQCC*	2431.90	<i>n.d.</i>	<i>n.d.</i>	(Zhang et al., 2006)
μ -CnIIIB	ZGCCGEPNLCFTRWCRNNARCCRQQ	2922.15	2922.11	2922.07	(Zhang et al., 2006)
μ -CnIIIC	ZGCCNGPKGCSKWCARDHARCC*	2373.88	2373.83	2373.86	(Benoit et al., 2008)
δ -CnVIA	YECYSTGTFCGINGGLCCSNLCLFFVC	3358.36	3358.31	<i>n.d.</i>	(Bulaj et al., 2001)
δ -CnVIB	DECFSGTFCGTFKOGGLCCSARCFSSFFCISLEF	3527.40	3527.39	<i>n.d.</i>	(Favreau, 1999)
δ -CnVIC	DECFSGTFCGTFKOGGLCCSARCFSSFFCISLEF	3505.46	3505.38	<i>n.d.</i>	(Favreau, 1999)
δ -CnVID	DECFSGTFCGTFKOGGLCCSARCFSSFFCISLEF	3539.44	3539.41	<i>n.d.</i>	(Favreau, 1999)
ω -CnVIA	CKGKGAOCTRLMYDCCHGSCSSSKGRC*	2846.14	2846.15	2846.13	(Favreau et al., 2001)
CcTx	AOWLVP <u>SQ</u> ITCCGYNOGTMCO S CMCTNTC	4115.51	4115.52	4115.58	(Le Gall F. et al., 1999)

Z, O, S and * represent post-translationally modified residues, respectively pyroglutamate, hydroxyproline, O-glycosylated serine and C-terminal amidation. *n.d.* stands for not detected.

Most surprisingly, only 73 common masses were found between milked and dissected venoms, thus representing only 48% of the milked venom components. Such a low percentage cannot solely be explained by intra-species variations, especially since 19 specimens were dissected, while the 67 milkings were performed over 3 months with 6 specimens, and all came from the same colony and were collected within 24 hours.

Therefore, it is tempting to believe that the additional compounds found in the milked venom are produced by other organs than the venom duct solely. In support of this hypothesis, a recent study describes, using molecular cloning techniques, several typical conopeptide sequences specifically expressed in the salivary glands of *Conus pulicarius* (Biggs, Olivera, & Kantor, 2008). Indeed, the salivary glands are connected to the radular sac, implying that peptides produced in this organ can be released in the milked venom only. The saliva may further play a role to dissolve the venom that is known to be very viscous inside the duct, allowing it to be injected through the harpoon. Other tissues may also participate in the production of the injected venom, including the snout glands and accessory salivary glands. In addition, a study describing the anatomy of *Conus californicus* also proposed that components may be produced in epithelial zones other than venom duct tissues, such as the connecting tissue between the venom duct and the pharynx (Marshall et al., 2002). Clearly, our results show that to obtain the most exhaustive mass list of venom components existing within one *Conus* species, both dissected and milked venom pools should be studied thoroughly using complementary bioanalytical techniques. Indeed, in this study, more than 1700 molecular masses were detected, representing more than ten times the number of venom components usually cited.

3.5 Concluding remarks

At times where natural component sources should be preserved, it is important to maximize the valuable information that one can obtain from a single sample. The different types of sample preparations and instruments that are available in the field of mass spectrometry are now quite vast. Focusing on one single technique may unfortunately lead to an incomplete investigation resulting in a waste of invaluable material that would have a lot more information to offer. By combining analyses of widely used MS instruments such as MALDI-MS and ESI-MS, we demonstrated, using the case of one cone snail species, to what extent

one can yield much larger data sets from one single sample available in tiny microgram amounts. Of course, the time and sample consumption, as well as the cost of such techniques must be pondered, but other MS instrumentation may in some cases allow to obtain even more out of a sample, as was illustrated recently using Fourier Transform mass spectrometry (Quinton, Le Caer, Vinh, Gilles, & Chamot-Rooke, 2006). In our case, the relatively low overlap between both techniques will thus provide us with a much broader spectrum of information to analyze than if only one preferred technique was used. In addition, the overlapped masses do not represent the most prominent ones, reinforcing the idea that important components may be left aside using solely one technique. It is also interesting to note that the hydrophobicity properties of components are not revealed in the same way depending on the MS technique employed. Finally, as it has been previously reported, but not to the same extent, we also demonstrated that milked venom provides highly valuable complementary information that adds to that obtained from dissected venom, giving access to biomolecules that could not be identified by other means. While demonstrating that cone snail venom complexity appears largely underestimated, this work constitutes the first step to a full venom proteome.

Acknowledgements

This project is part of the “Venomics” genome project initiated by the International Society of Toxinology dedicated to the understanding of the function and evolution of venomous systems in various phyla (Menez et al., 2006). This study has been performed as part of the CONCO cone snail genome project for health (<http://www.conco.eu>) within the 6th Framework Program (LIFESCIHEALTH-6 Integrated Project LSHB-CT-2007, contract number 037592). We are most grateful to the European Commission for funding. We would like to express our deepest gratitude to the Government of New Caledonia, the French Navy, Fabrice Colin (IRD-Nouméa), Napoléon Colombani (IRD-Nouméa), Claude Payri (IRD-Nouméa) and

Jacques Pusset (Toxinomics Foundation office in Nouméa) for their constant support. We also want to thank Frederic Perret, Célia Boiteau and Olivier Cheneval from our laboratories for their advice on the different MS preparation methods and HPLC settings. This article is dedicated to the memory of André Ménez without whom none of this would have been possible.

References

- Apffel, A., Fischer, S., Goldberg, G., Goodley, P. C., & Kuhlmann, F. E. (1995). Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases. *J.Chromatogr.A*, 712, 177-190.
- Benoit, E., Favreau, P., Schlumberger, S., Cordova, MA., Tytgat, J., Stocklin, R. et al. (2008). A new mu-conotoxin from *Conus* consors that atypically targets sodium channels in unmyelinated and myelinated nerve fibers. In Abstract Book (Ed.), *16th European Section Meeting of the International Society on Toxinology* (.).
- Biggs, J. S., Olivera, B. M., & Kantor, Y. I. (2008). Alpha-conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon*, 52, 101-105.
- Bingham, J. P., Jones, A., Lewis, R. J., Andrew, P. R., & Alewood, D. (1996). *Conus* venom peptides (conopeptides): inter-species, intra-species and within individual variation revealed by ionspray mass spectrometry. In E.Lazarovici, M. E. Spira, & E. Zlotkin (Eds.), *Biomedical Aspects of Marine Pharmacology* (pp. 13-27). Fort Collins: Alaken.
- Bulaj, G., DeLaCruz, R., zimi-Zonooz, A., West, P., Watkins, M., Yoshikami, D. et al. (2001). Delta-conotoxin structure/function through a cladistic analysis. *Biochemistry*, 40, 13201-13208.

- Calvete, J. J., Escolano, J., & Sanz, L. (2007). Snake venomomics of Bitis species reveals large intragenus venom toxin composition variation: application to taxonomy of congeneric taxa. *J.Proteome.Res.*, 6, 2732-2745.
- Creer, S., Malhotra, A., Thorpe, R. S., Stocklin, R. S., Favreau, P. S., & Hao Chou, W. S. (2003). Genetic and ecological correlates of intraspecific variation in pitviper venom composition detected using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS) and isoelectric focusing. *J.Mol.Evol.*, 56, 317-329.
- Escoubas, P., Celerier, M. L., & Nakajima, T. (1997). High-performance liquid chromatography matrix-assisted laser desorption/ionization time-of-flight mass spectrometry peptide fingerprinting of tarantula venoms in the genus Brachypelma: chemotaxonomic and biochemical applications. *Rapid Commun.Mass Spectrom.*, 11, 1891-1899.
- Escoubas, P., Sollod, B., & King, G. F. (2006). Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon*, 47, 650-663.
- Favreau, P. (1999). *Physico-chemical and pharmacological study of new toxins from the piscivorous cone snail Conus Consors.*
- Favreau, P., Gilles, N., Lamthanh, H., Bournaud, R., Shimahara, T., Bouet, F. et al. (2001). A new omega-conotoxin that targets N-type voltage-sensitive calcium channels with unusual specificity. *Biochemistry*, 40, 14567-14575.
- Favreau, P., Krimm, I., Le, G. F., Bobenrieth, M. J., Lamthanh, H., Bouet, F. et al. (1999). Biochemical characterization and nuclear magnetic resonance structure of novel alpha-conotoxins isolated from the venom of Conus consors. *Biochemistry*, 38, 6317-6326.

- Favreau, P., Menin, L., Michalet, S., Perret, F., Cheneval, O., Stocklin, M. et al. (2006). Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen. *Toxicon*, 47, 676-687.
- Garcia, M. C. (2005). The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.*, 825, 111-123.
- Garrett, J. E., Buczek, O., Watkins, M., Olivera, B. M., & Bulaj, G. (2005). Biochemical and gene expression analyses of conotoxins in *Conus textile* venom ducts. *Biochem.Biophys.Res.Comm.*, 328, 362-367.
- Gobom, J., Schuerenberg, M., Mueller, M., Theiss, D., Lehrach, H., & Nordhoff, E. (2001). Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal.Chem.*, 73, 434-438.
- Gowd, K. H., Dewan, K. K., Iengar, P., Krishnan, K. S., & Balaram, P. (2008). Probing peptide libraries from *Conus achatinus* using mass spectrometry and cDNA sequencing: identification of delta and omega-conotoxins. *J.Mass Spectrom.*, 43, 791-805.
- Herzig, V. & Hodgson, W. C. (2008). Intersexual variations in the pharmacological properties of *Coremiocnemis tropix* (Araneae, Theraphosidae) spider venom. *Toxicon*.
- Hopkins, C., Grilley, M., Miller, C., Shon, K. J., Cruz, L. J., Gray, W. R. et al. (1995). A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J.Biol.Chem.*, 270, 22361-22367.

- Jakubowski, J. A., Kelley, W. P., Sweedler, J. V., Gilly, W. F., & Schulz, J. R. (2005). Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J.Exp.Biol.*, *208*, 2873-2883.
- Kelley, W. P., Schulz, J. R., Jakubowski, J. A., Gilly, W. F., & Sweedler, J. V. (2006). Two toxins from *Conus striatus* that individually induce tetanic paralysis. *Biochemistry*, *45*, 14212-14222.
- Le Gall F., Favreau, P., Benoit, E., Mattei, C., Bouet, F., Menou, J. L. et al. (1999). A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na⁺-dependent mechanism. *Eur.J.Neurosci.*, *11*, 3134-3142.
- Lewis, R. J. & Garcia, M. L. (2003). Therapeutic potential of venom peptides. *Nat.Rev.Drug Discov.*, *2*, 790-802.
- Marshall, J., Kelley, W. P., Rubakhin, S. S., Bingham, J. P., Sweedler, J. V., & Gilly, W. F. (2002). Anatomical correlates of venom production in *Conus californicus*. *Biol.Bull.*, *203*, 27-41.
- Martinez, J. S., Olivera, B. M., Gray, W. R., Craig, A. G., Groebe, D. R., Abramson, S. N. et al. (1995). alpha-Conotoxin EI, a new nicotinic acetylcholine receptor antagonist with novel selectivity. *Biochemistry*, *34*, 14519-14526.
- Menez, A., Stocklin, R., & Mebs, D. (2006). 'Venomics' or : The venomous systems genome project. *Toxicon*, *47*, 255-259.
- Newton, K. A., Clench, M. R., Deshmukh, R., Jeyaseelan, K., & Strong, P. N. (2007). Mass fingerprinting of toxic fractions from the venom of the Indian red scorpion,

- Mesobuthus tamulus: biotope-specific variation in the expression of venom peptides. *Rapid Commun.Mass Spectrom.*, 21, 3467-3476.
- Olivera, B. M. (1997). E.E. Just Lecture, 1996. Conus venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol.Biol.Cell*, 8, 2101-2109.
- Olivera, B. M., Gray, W. R., Zeikus, R., McIntosh, J. M., Varga, J., Rivier, J. et al. (1985). Peptide neurotoxins from fish-hunting cone snails. *Science*, 230, 1338-1343.
- Padliya, N. D. & Wood, T. D. (2008). Improved peptide mass fingerprinting matches via optimized sample preparation in MALDI mass spectrometry. *Anal.Chim.Acta*, 627, 162-168.
- Perchuc, A. M., Menin, L., Stocklin, R., Buhler, B., & Schoni, R. (2005). The potential of Bothrops moojeni venom in the field of hemostasis. Established use and new insights. *Pathophysiol.Haemost.Thromb.*, 34, 241-245.
- Pimenta, A. M., Stocklin, R., Favreau, P., Bougis, P. E., & Martin-Eauclaire, M. F. (2001). Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of Tityus serrulatus (Scorpiones, Buthidae). *Rapid Commun.Mass Spectrom.*, 15, 1562-1572.
- Quinton, L., Le Caer, J. P., Vinh, J., Gilles, N., & Chamot-Rooke, J. (2006). Fourier transform mass spectrometry: a powerful tool for toxin analysis. *Toxicon*, 47, 715-726.
- Ramilo, C. A., Zafaralla, G. C., Nadasdi, L., Hammerland, L. G., Yoshikami, D., Gray, W. R. et al. (1992). Novel alpha- and omega-conotoxins from Conus striatus venom. *Biochemistry*, 31, 9919-9926.

- Sanz, L., Escolano, J., Ferretti, M., Biscoglio, M. J., Rivera, E., Crescenti, E. J. et al. (2008). Snake venomomics of the South and Central American Bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered from proteomic versus transcriptomic analysis. *J.Proteomics.*, 71, 46-60.
- Schroeder, C. I., Ekberg, J., Nielsen, K. J., Adams, D., Loughnan, M. L., Thomas, L. et al. (2008). Neuronally selective mu -conotoxins from *Conus striatus* utilise an alpha - helical motif to target mammalian sodium channels. *J.Biol.Chem.*.
- Schuerenberg, M., Luebbert, C., Eickhoff, H., Kalkum, M., Lehrach, H., & Nordhoff, E. (2000). Prestructured MALDI-MS sample supports. *Anal.Chem.*, 72, 3436-3442.
- Schwartz, E. F., Camargos, T. S., Zamudio, F. Z., Silva, L. P., Bloch, C., Jr., Caixeta, F. et al. (2008). Mass spectrometry analysis, amino acid sequence and biological activity of venom components from the Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon*, 51, 1499-1508.
- Stocklin, R. & Favreau, P. (2002). Proteomics of venom peptides. In A.Menez (Ed.), *Perspectives in molecular toxinology* (pp. 107-123). Chichester, UK: John Wiley & Sons Ltd.
- Teichert, R. W., Rivier, J., Dykert, J., Cervini, L., Gulyas, J., Bulaj, G. et al. (2004). AlphaA-Conotoxin OIVA defines a new alphaA-conotoxin subfamily of nicotinic acetylcholine receptor inhibitors. *Toxicon*, 44, 207-214.
- Zhang, M. M., Fiedler, B., Green, B. R., Catlin, P., Watkins, M., Garrett, J. E. et al. (2006). Structural and functional diversities among mu-conotoxins targeting TTX-resistant sodium channels. *Biochemistry*, 45, 3723-3732.

4.4 Concluding remarks

The high-resolution chromatographic separation of two venom samples (IV and DV) in combination with two different MS techniques - ESI-MS and MALDI-MS - enabled more than 1,700 different components to be found. The results also highlighted the variability in the mass range distribution and the relative hydrophobicity between the two datasets. It also clearly proved that both MS techniques were complementary as only 20% of the masses found overlapped. The total number of components was more than ten times that usually found hitherto in cone snail venom content. The 10-fold increase was further confirmed by other studies on other cone snail venoms^{34,35}. This work demonstrated that the complexity of cone snail venoms in general was clearly underestimated.

4.5 Reference List

1. Adams A. *Narrative of the voyage of H.M.S. Samarang*; - ed.; 1848; Vol. -; pp. -.
2. Flecker, H. Cone Shell Mollusc Poisoning, With Report of a Fatal Case. *The Medical Journal of Australia* 1936, 1, 464-466.
3. Hermitte, L. C. D. Venomous Marine Molluscs of the Genus *Conus*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1946, 39 (6), 485-501.
4. Kohn, A. J. Cone shell stings; recent cases of human injury due to venomous marine snails of the genus *Conus*. *Hawaii Med. J.* 1958, 17 (6), 528-532.
5. Kline, G. F. Notes on the stinging operation of *Conus*. *Nautilus* 1956, 69 (3), 76-78.
6. Kohn, A. J.; Saunders, P. R.; WIENER, S. Preliminary studies on the venom of the marine snail *Conus*. *Ann. N. Y. Acad. Sci.* 1960, 90 (-), 706-725.
7. Endean, R.; Rudkin, C. Further studies of the venoms of Conidae. *Toxicon* 1965, 2 (4), 225-249.
8. Endean, R.; Rudkin, C. Studies of the venoms of some Conidae. *Toxicon* 1963, 1 (2), 49-64.
9. Whysner, J. A.; Saunders, P. R. Purification of the lethal fraction of the venom of the marine snail *Conus californicus*. *Toxicon* 1966, 4 (3), 177-181.
10. Cruz, L. J.; Corpuz, G.; Olivera, B. M. A preliminary study of *Conus* venom protein. *The Veliger* 1975, 18 (3), 302-308.
11. Endean, R.; Parish, G.; Gyr, P. Pharmacology of the venom of *Conus geographus*. *Toxicon* 1974, 12 (2), 131-138.
12. Endean, R.; Gyr, P.; Parish, G. Pharmacology of the venom of the gastropod *Conus magus*. *Toxicon* 1974, 12 (2), 117-129.
13. Endean, R.; Williams, H.; Gyr, P.; Surridge, J. Some effects on muscle and nerve of crude venom from the gastropod *Conus striatus*. *Toxicon* 1976, 14 (4), 267-274.
14. Endean, R.; Gyr, P.; Surridge, J. The pharmacological actions on guinea-pig ileum of crude venoms from the marine gastropods *Conus striatus* and *Conus magus*. *Toxicon* 1977, 15 (4), 327-337.

15. Endean, R.; Surridge, J.; Gyr, P. Some effects of crude venom from the cones *Conus striatus* and *Conus magus* on isolated guinea-pig atria. *Toxicon* 1977, 15 (5), 369-374.
16. Endean, R.; Gyr, P.; Surridge, J. The effects of crude venoms of *Conus magus* and *Conus striatus* on the contractile response and electrical activity of guinea-pig cardiac musculature. *Toxicon* 1979, 17 (4), 381-395.
17. Cruz, L. J.; Gray, W. R.; Olivera, B. M. Purification and properties of a myotoxin from *Conus geographus* venom. *Arch. Biochem. Biophys.* 1978, 190 (2), 539-548.
18. Kobayashi, J.; Ohizumi, Y.; Nakamura, H.; Hirata, Y. Pharmacological study on the venom of the marine snail *Conus textile*. *Toxicon* 1981, 19 (6), 757-762.
19. Kobayashi, J.; Nakamura, H.; Hirata, Y.; Ohizumi, Y. Isolation of a cardiotonic glycoprotein, striatoxin, from the venom of the marine snail *Conus Striatus*. *Biochem. Biophys. Res. Commun.* 1982, 105 (4), 1389-1395.
20. Kobayashi, J.; Nakamura, H.; Ohizumi, Y. Excitatory and inhibitory effects of a myotoxin from *Conus magus* venom on the mouse diaphragm, the guinea-pig atria, taenia caeci, ileum and vas deferens. *Eur. J. Pharmacol.* 1982, 86 (2), 283-286.
21. McIntosh, J. M.; Olivera, B. M.; Cruz, L. J.; Gray, W. R. Gamma-carboxyglutamate in a neuroactive toxin. *J. Biol. Chem.* 1984, 259 (23), 14343-14346.
22. Jimenez, E. C.; Olivera, B. M.; Cruz, L. J. Localization of enzymes and possible toxin precursors in granules from *Conus striatus* venom. *Toxicon* 1983, 21 (Suppl.3), 199-202.
23. Olivera, B. M.; Gray, W. R.; Zeikus, R.; McIntosh, J. M.; Varga, J.; Rivier, J.; de, S., V; Cruz, L. J. Peptide neurotoxins from fish-hunting cone snails. *Science* 1985, 230 (4732), 1338-1343.
24. Krishnamurthy, T.; Prabhakaran, M.; Long, S. R. Mass spectrometric investigations on *Conus* peptides. *Toxicon* 1996, 34 (11-12), 1345-1359.
25. Nakamura, T.; Yu, Z.; Fainzilber, M.; Burlingame, A. L. Mass spectrometric-based revision of the structure of a cysteine-rich peptide toxin with gamma-carboxyglutamic acid, TxVIIA, from the sea snail, *Conus textile*. *Protein Sci.* 1996, 5 (3), 524-530.
26. Craig, A. G.; Jimenez, E. C.; Dykert, J.; Nielsen, D. B.; Gulyas, J.; Abogadie, F. C.; Porter, J.; Rivier, J. E.; Cruz, L. J.; Olivera, B. M.; McIntosh, J. M. A novel post-translational modification involving bromination of tryptophan. Identification of the residue, L-6-bromotryptophan, in peptides from *Conus imperialis* and *Conus radiatus* venom. *J. Biol. Chem.* 1997, 272 (8), 4689-4698.

27. Stocklin, R.; Favreau, P. Proteomics of venom peptides. In *Perspectives in molecular toxinology*, - ed.; Menez, A., Ed.; John Wiley & Sons Ltd: Chichester, UK, 2002; pp 107-123.
28. Jakubowski, J. A.; Sweedler, J. V. Sequencing and mass profiling highly modified conotoxins using global reduction/alkylation followed by mass spectrometry. *Anal. Chem.* 2004, 76 (22), 6541-6547.
29. Jakubowski, J. A.; Kelley, W. P.; Sweedler, J. V.; Gilly, W. F.; Schulz, J. R. Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J. Exp. Biol.* 2005, 208 (Pt 15), 2873-2883.
30. Jakubowski, J. A.; Kelley, W. P.; Sweedler, J. V. Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon* 2006, 47 (6), 688-699.
31. Vianna Braga, M. C.; Konno, K.; Portaro, F. C.; de Freitas, J. C.; Yamane, T.; Olivera, B. M.; Pimenta, D. C. Mass spectrometric and high performance liquid chromatography profiling of the venom of the Brazilian vermivorous mollusk *Conus regius*: feeding behavior and identification of one novel conotoxin. *Toxicon* 2005, 45 (1), 113-122.
32. Quinton, L.; Le Caer, J. P.; Vinh, J.; Gilles, N.; Chamot-Rooke, J. Fourier transform mass spectrometry: a powerful tool for toxin analysis. *Toxicon* 2006, 47 (6), 715-726.
33. Menez, A.; Stocklin, R.; Mebs, D. 'Venomics' or : The venomous systems genome project. *Toxicon* 2006, 47 (3), 255-259.
34. Davis, J.; Jones, A.; Lewis, R. J. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 2009, 30 (7), 1222-1227.
35. Dutertre, S.; Jin, A. H.; Kaas, Q.; Jones, A.; Alewood, P. F.; Lewis, R. J. Deep venomomics reveals the mechanism for expanded Peptide diversity in cone snail venom. *Mol. Cell Proteomics.* 2013, 12 (2), 312-329.

5. Transcriptomics

5.1 Introduction

As mentioned previously, the latest advances in chromatography, MS techniques and the combination of these two have generated an enormous amount of data. The extraction of valuable information from this huge mass has thus become the current major challenge. Hitherto, in the proteomic field, the characterisation of peptides or proteins was mostly carried out by *de novo* MS/MS sequencing. Unfortunately, this tends to be more easily said than done: indeed, this method is very slow and tedious as it can take up to several weeks to fully characterize one peptide sequence.

With the advent of new science disciplines like bioinformatics and their databases, such as the well known Swissprot database created in the late '90s ¹, the characterisation of known peptide sequences became easier, but at the time was still posing considerable problems in terms of revealing new peptide sequences.

Later, with the combination of molecular biology techniques and the now retroactively called first-generation sequencing techniques ², sequence information on targeted peptides and proteins became available and were implemented in conjunction with databases, which increased the number of possible characterisations. Unfortunately, the data available in such databases as a means of comparison with experimental proteomic data was still insufficient to reveal the novelty and magnitude of a proteome. It is only the arrival of second-generation sequencing techniques that enabled the amount of available sequences to grow exponentially and with it, the possibilities of comparison with proteomic data.

5.2 The beginning: cDNA cloning

In the field of venomomics, and more specifically of cone snail venom, the use of transcripts directly derived from the venom gland tissue to assist the characterisation or to confirm the proper MS/MS characterisation of cone snail venom peptides dates back to the very late '80s³. At that time, the technique consisted of creating a cDNA library of poly(A+) RNA with mixed oligonucleotide probes based on conserved regions of the propeptide transcripts. Relevant clones were then selected and used to obtain the full corresponding propeptide transcripts. These were then sequenced using traditional first-generation sequencing methods. Not only did this bring about the advantage of rapid confirmation of MS/MS sequences – compared with *de novo* MS/MS sequencing – but it also allowed other members of the selected peptide family to be revealed (or found). The discovery of homologous sequences was also a stepping stone into understanding the structure of the propeptide transcripts and their impact on the variability of the venom content⁴. Furthermore, these homologous sequences would also come and grow the ranks of peptide databases⁵. Unfortunately, the most serious drawback of this technique was that it kept focusing on already partially or fully characterized peptides or peptide families. Indeed, the probes were built each time specifically to pick out the transcripts of a given interest, which therefore limited the discovery of new components and families.

As a complement to the characterisation by mass spectrometry, the introduction of molecular biology in the field of venomomics not only, appeared to be really interesting for the global knowledge of the evolution of cone snail species but also more specifically in the venom processing⁶. Although the vastness of the peptide diversity in cone snail venom transcriptomes was suspected, and the same scenario with proteomics later, it is the scale of this variability that was underestimated. Indeed, most peptides differed from one species to another and multiplying this by the number of species involved would mean that using cDNA

cloning for all peptides would be far too time consuming, not to mention the risk of missing new peptide families on the way.

Finally the mid-2000s came and with them the first global molecular studies of cone snail venom glands using of ESTs (Expressed Sequences Tags) ^{7,8}. Although, these analyses were achieved with first-generation sequencers, the output was quite substantial with 429 ESTs (of which 221 were toxin sequences) from piscivorous *Conus striatus* venom ducts and 897 ESTs (with nearly half of them representing conopeptide sequences) from *Conus litteratus* venom ducts. The results of these studies ^{7,8}, provided by the same group, constituted the first comprehensive sets of cone snail gene sequences. Their partially sequenced ESTs enabled the identification of the most common known conopeptide families – as well as some new ones. Unfortunately, these studies only revealed a fraction of the potential content of a cone snail venom duct transcriptome, due to some hindering factors affecting the methods and techniques then available (like the cDNA cloning used and their random selection, the DNA sequencing technique itself or the errors generated by the PCR amplifications). Thanks to these instruments and the Human Genome Project, the way to the next-generation sequencing was being paved.

5.3 Next-generation sequencing (NGS)

The first two NGS transcriptomes appeared in 2011. They both used the same sequencing technique, the 454 pyrosequencing of ESTs ^{9,10}. Although both transcriptome reports provide an unprecedented amount of data that had to be analysed due to the huge amount of reads provided by the second generation sequencing techniques, their results and discussions take a different slant. Whereas Hu et al. focus more on the methods and techniques used for their transcriptome and their partial genome, we here provide (see article below) a more

exhaustive analysis of the transcriptome content with, for instance, Gene Ontology annotations and a detailed profiling of the conopeptide transcripts.

High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*

Yves Terrat¹, Daniel Biass³, Sébastien Dutertre³, Philippe Favreau³, Mairo Remm⁴, Reto Stöcklin³, David Piquemal², Frédéric Ducancel¹

¹ Laboratoire d'Ingénierie des Anticorps pour la Santé (LIAS), CEA/DSV/iBiTec-s/SPI, CEA Saclay, Bt. 152, 91191 Gif-sur-Yvette Cedex, France;

² Skuldtech, 134, rue du Curat, Bat. Amarante, 34090 Montpellier, France;

³ Atheris Laboratories, CH-1233 Bernex, Geneva, Switzerland;

⁴ Estonian Biocentre, Riia str. 23, 51010 Tartu, Estonia.

Publication data

Published the 06.11.2011 in Toxicon, Vol. 59, Issue 1, p.34-46.

Pubmed ID: 22079299

DOI: 10.1016/j.toxicon.2011.10.001

Cited 37 times on Google Scholar, 30 times on ScienceDirect.

Keywords

Toxin, conopeptide, cone snail, *Conus consors*, pyrosequencing, transcriptome

Personal contribution

Part of the data analysis of the transcriptome, part of the manuscript writing and proofreading.

Abstract

Although cone snail venoms have been intensively investigated in the past few decades, little is known about the whole conopeptide and protein content in venom ducts, especially at the transcriptomic level. If most of the previous studies focusing on a limited number of sequences have contributed to a better understanding of conopeptide superfamilies, they did not give access to a complete panorama of a whole venom duct. Additionally, rare transcripts were usually not identified due to sampling effect. This work presents the data and analysis of a large number of sequences obtained from high throughput⁴⁵⁴ sequencing technology using venom ducts of *Conus consors*, an Indo-Pacific living piscivorous cone snail. A total of 213,561 Expressed Sequence Tags (ESTs) with an average read length of 218 base pairs (bp) have been obtained. These reads were assembled into 65,536 contiguous DNA sequences (contigs) then into 5039 clusters. The data revealed 11 conopeptide superfamilies representing a total of 53 new isoforms (full length or nearly full-length sequences). Considerable isoform diversity and major differences in transcription level could be noted between superfamilies. A, O and M superfamilies are the most diverse. The A family isoforms account for more than 70% of the conopeptide cocktail (considering all ESTs before clustering step). In addition to traditional superfamilies and families, minor transcripts including both cysteine free and cysteine-rich peptides could be detected, some of them figuring new clades of conopeptides. Finally, several sets of transcripts corresponding to proteins commonly recruited in venom function could be identified for the first time in cone snail venom duct. This work provides one of the first large-scale EST project for a cone snail venom duct using next-generation sequencing, allowing a detailed overview of the venom duct transcripts. This leads to an expanded definition of the overall cone snail venom duct transcriptomic activity, which goes beyond the cysteine-rich conopeptides. For instance, this study enabled to detect proteins involved in common post-translational maturation and folding, and to reveal compounds classically involved in hemolysis and mechanical penetration of the venom into the prey. Further comparison with proteomic and genomic data will lead to a better understanding of conopeptides diversity and the underlying mechanisms involved in conopeptide evolution.

Abbreviations:

BLAST, Basic Local Alignment Search Tool; bp, base pair; CCG, Common Cellular Gene; EST, Expressed Sequence Tag; GO, Gene Ontology; LTR, Long Terminal Repeat; TE, Transposable Element; TRG, Toxin Related Gene; XB, Xenobiotic.

1. Introduction

Cone snails of the genus *Conus* are predatory marine gastropods that utilize venom to capture prey. *Conus* venoms mainly consist of small disulfide-rich peptides commonly referred to as conotoxins or conopeptides. Each of the near 700 *Conus* species synthesizes its own characteristic repertoire of toxic peptides. It has been estimated that the toxin repertoire of cone snails comprises more than 100,000 different bioactive compounds with various neurological targets (Terlau and Olivera, 2004). The venom cocktail of cone snails is further extended through the addition of post-translational modifications that increase toxin potency (Pisarewicz et al., 2005; Lopez-Vera et al., 2008) and assist the stabilization of the three-dimensional molecular structure (Craig et al., 2001). These components target a variety of ion channels, transporters and receptors besides the interest in their natural functions in venoms. They are therefore of much interest as drug leads (Favreau and Stöcklin, 2009). Even though cone snail venoms have been assiduously studied during the past few decades, a comprehensive study of the transcriptomic content of a cone snail venom duct is a new step in this field. Indeed, earlier studies aiming at specific conopeptide superfamilies and using only a restrained number of sequences did not lead to a complete overview of a whole venom composition. Furthermore, sampling effects have limited the identification of rare transcripts.

A number of research groups have sequenced portions of cone snail venom duct transcriptomes to identify the genes contributing to venoms. Most of these previous studies

have relied on cloning of cDNA libraries and Sanger sequencing, generating important, but ultimately limited, data. More than half of the expressed sequence tags (ESTs) from these studies have, in most cases, been found to code for toxin genes (Pi et al., 2006a,b) and a large proportion of the remaining ESTs to code for genes involved in transcription, translation, cell regulation and metabolism. The application of next-generation sequencing, though not without its own drawbacks, should alleviate issues of low coverage and provide a more complete characterization of the genes contributing to cone snail venoms. The current transcriptomic analysis is integrated in the CONCO project (<http://www.conco.eu>). This project focuses on the discovery and the development of new therapeutically relevant molecules issued from the venomous marine cone snail species *Conus consors*. Through the deep and exhaustive investigation of the animal biodiversity, of its genome, its venom gland transcriptome and its venom proteome, CONCO aims at exploiting in a sustainable way the great richness offered by these animals to discover the drugs of tomorrow (Dutertre et al., 2010; Kaufenstein et al., 2011). The present manuscript presents analysis of data generated by high throughput sequencing. Not only this work focuses on characterizing the most abundant toxin-encoding transcripts in the venom-gland transcriptome, but it provides also analyses of the remaining part of the transcriptome, presenting new insights on the venom cocktail used by the cone snail to capture his prey.

2. Material and methods

2.1 Species collection and RNA extraction

Specimens of *C. consors* were collected in the Chesterfield Islands (New Caledonia) in the frame of the CONFIELD scientific expedition in June 2007. Venom ducts were dissected from three living snails, and were stored in RNA-later stabilization reagent following manufacturer's instructions (Qiagen). Total RNA was extracted with TRIzol reagent using

standard protocols (Invitrogen). The cDNA library was not subjected to a normalization step prior to sequencing.

2.2. 454 Pyrosequencing and contig assembly

Approximately 1 mg of the adaptator-ligated cDNA population was sheared by nebulization and DNA sequencing was performed following protocols for the genome Sequencer GS FLX System (Roche diagnostic). Sequence reads were processed in order to exclude low quality and poly A+ tracts using Trace2dbEST (Parkinson and Blaxter, 2004). Subsequently, assembly was undertaken with SeqMan pro (DNASTAR, USA) using high stringency clustering parameters (100% identity between reads with 40 nucleotides sequence overlap). The contigs were then assembled into clusters using Cd-hit software (Huang et al., 2010) with a 80% identity threshold. Each cluster is represented by its longest sequence for further similarity-based annotation.

2.3. Annotation of ESTs sequences

Bioinformatic processing was carried out using a combination of public softwares and home-made scripts using PERL programming language. The EST workflow is summarized in Figure 18. First, representative sequences of each cluster were BLAST-annotated (Altschul et al., 1990) against UniProt/Swiss-Prot (v56.2), Uniprot/TrEMBL (v39.2) (UniProt Consortium, 2011), in-house toxin database (named ToxReIDB) and Repbase repository dedicated to transposable elements (Jurka et al., 2005). Most toxin-related sequences from cone snails are conopeptides but numerous proteins have been convergently recruited into the venoms of various animals (Fry et al., 2009). For this reason, toxin sequences from a wide range of taxonomic groups are integrated in our regularly updated home-made toxin database to increase new toxin detection. A combination of BLASTn and BLASTx algorithms was used for contig annotation. This later was considered successful when the best match

displayed an expected e-value $\sim 1.10^{-3}$. After the first round of data treatment, sequences were parsed using a home-made PERL script to classify representative sequences in 6 categories: Common cellular genes (CCGs), Conopeptides, Toxin-related genes (TRGs), Xenobiotics (XBs), Transposable elements (TEs) and Unknown sequences (UKs).

Sequences related to a given cluster and encoding for a putative conopeptide family were submitted to a home-made program for isoforms identification. To increase sensitivity of BLAST searches to detect conopeptides, a second BLAST step was performed using as queries the first dataset of *C. consors* sequences identified so far.

Since many venom genes have been shown to have evolved from genes coding for normal cellular products, it is therefore difficult to reject an expressed gene as a venom component just because it has some similarity with a gene coding for a protein involved in a “conventional” cellular process. Due to this issue, an estimation of the transcription level (see further) for the different contigs was performed and the top-ranked hits that were not identified as conopeptides, XBs, TEs, or TRGs were retrieved. These were then searched for a putative toxin role in the literature.

For naming the different conopeptide superfamilies, the following nomenclature was used: a three letters code designates the species (Cco for *C. consors*), a capital letter (A, O, M, P, S, T) or a full-length name (Conantokin, Con-opressin, Conantokin, Conodipin, Conkunitzin) indicates the superfamily or family attribution and finally the last lower case letter distinguishes isoforms. This nomenclature is different for two reasons from the classical nomenclature based on pharmacological properties and using two letters to name each species as implemented in the Conoserver database (Kaas et al., 2010). First, the expanding number of protein and nucleic sequences generated by high throughput analysis will necessarily lead to much confusion if only two letters are still used to name each species. Second, a similarity/phylogenetical nomenclature was chosen instead of the pharmaceutical

one because no pharmaceutical property could be predicted with high confidence from the primary sequence.

Xenobiotic sequences were first detected by key word searches of the first BLAST hit. As the closest BLAST hit is often not the nearest neighbor (Stanhope et al., 2001), phylogenetical analysis of suspected xenobiotics were performed using every BLAST hit associated with an e-value threshold below 1E-03 to confirm that such sequences were members of particular symbionts.

2.4. Gene ontology

The functional genomic study was performed using BLAST2GO software (Conesa and Götzel, 2008) with the NCBI non-redundant protein database (with a cut-off e-value of 1.10×10^{-3}) using the EST contigs. For each contig, the GI accession (NCBI) of the best hits retrieved and the Gene Ontology (GO) accessions were mapped to GO terms according to molecular function, biological process and cellular component ontologies at a level which gives the most abundant category numbers (Ashburner et al., 2000).

2.5. Alignment and phylogeny analysis

DNA or amino acid (aa) sequence alignments were performed using Muscle v4 (Edgar, 2004) followed by manual adjustment and manually checked with the software Jalview (Waterhouse et al., 2009) or Bioedit (Ibis Therapeutics, Carlsbad, CA, USA). Prior to phylogenetical analysis, divergent and ambiguously aligned blocks were removed using Gblocks software (Talavera and Castresena, 2007). To choose the best-fitted model, JModelTest (Posada, 2009) and ProtTest (Abascal et al., 2005) were used. Phylogenetical analyses were carried out using maximum likelihood method with the PhyML software

(Guindon et al., 2009) and trees were visualized and annotated using the tree viewer of MEGA 4 (Tamura et al., 2007). Bootstrap support values based on 100 replicates.

2.6. Transcriptional profiling of conopeptides

Since EST contigs were created from a non-normalized library, they are supposed to be reflective of the population of RNA sequences in the originating tissue. More-over, the length of the different conopeptide families fall in the same length range (mean value for *C. consors*: 76.7, standard error: 12.3) which means that there was a minimum bias in the sampling sequencing procedure. For each isoform, a MegaBLAST search (Zhang et al., 2000) was performed on the raw data (non-assembled EST reads). Then, hits associated with a minimal e-value of 1E-10 were selected. In order to be sure that the multiple BLASTn hits given were not counted several times between the different queries, these were excluded from the analysis by a simple comparison of a given query hit to the complete hits dataset. Thus, the number of different hits from the raw data could be a relative estimation of the transcription level for a given isoform, including single base errors, small insertions or deletions (indels) due to pyrosequencing technology (Parameswaran et al., 2007).

3. Results

3.1. Sequencing and assembly of 454 pyrosequencing ESTs

The sequencing effort comprised a grand total of 46.5 mega base (Mb) making up 213,561 reads with a minimal size of 50 bp and average of 218 bp. Primer removal of raw sequences and subsequent contigs assembly resulted in 65,536 contigs averaging 240 bp. Singletons were excluded from the further analysis. Contigs clustered with Cd-hit-est lead to 5039 clusters (80% DNA identity with a minimal coverage of 50 bp). Statistics about EST database and clustering step are represented in Figure 18.

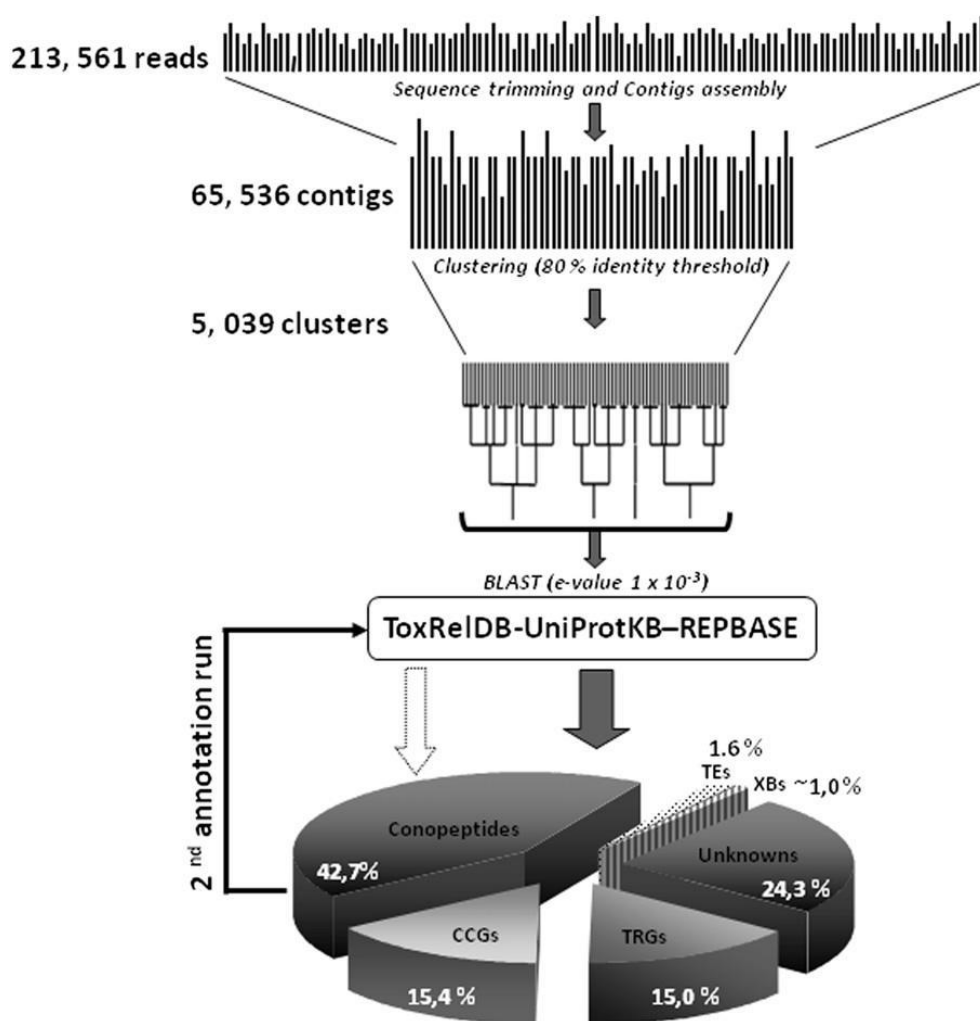


Figure 18: EST processing and description of the venom duct transcriptome contigs obtained by similarity searches.

The abbreviations used for the categories are: Common Cellular Genes (CCGs), Toxin-Related Genes (TRGs), Transposable Elements (TEs) and Xenobiotics (XBs).

3.2. Annotation of ESTs sequences, gene ontologies and species distribution of BLAST hits

Representative sequences of each cluster were analyzed with BLASTx and BLASTn programs. 48,556 contigs (74.09%) have significant similarities with proteins in the UniprotKB, Repbase or ToxRelDB databases; the remaining 16,980 contigs (25.91%) had no function assigned. The main part of identified sequences (28,003 contigs or 42.73%)

represent conopeptide sequences, most of them belonging to the A, O and M superfamilies. Nearly half of all contigs identified are similar to previously characterized conopeptides which is in agreement with previous classical cDNA cloning analysis (Pi et al., 2006a, 2006b). Common cellular genes represent 10,089 contigs (15.40%). It is noteworthy to mention that a significant part of contigs is identified as transposable elements (1046 or 1.60%) and xenobiotics (628 or 0.96%). Results are summarized in Figure 18.

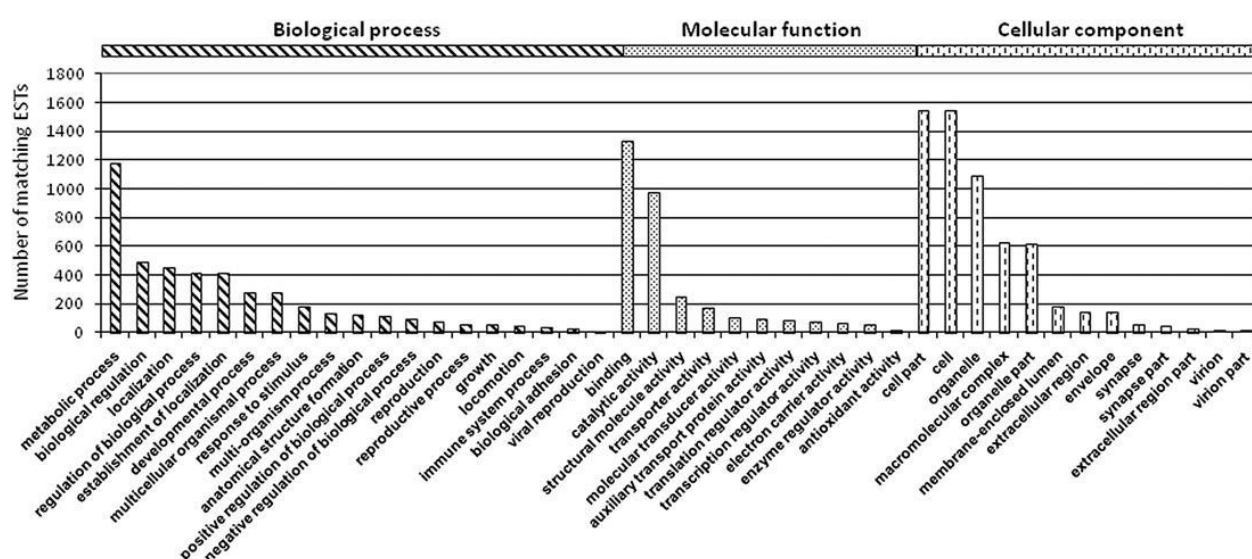


Figure 19. Gene ontology-sorted annotations.

Results are split into Biological process, Molecular function and Cellular component at the second level according to the standard gene ontology terms. The different ontology categories are represented on the X axis. Number of ESTs matching GO annotation terms (prior to clustering) is represented on the Y axis. As more than 80% of the contigs failed to be associated with any GO terms, this analysis represents annotation of the most conserved eukaryotic genes.

Gene ontology assignments were classified into Molecular Functions (MF), Biological Process (BP) and Cellular Components (CC) at the second level according to the standard gene ontology terms (<http://www.geneontology.org>). About 70% of all contigs failed to be associated with GO terms. This result is in agreement with previous data for most non-model species, especially for invertebrates (Vera et al., 2008). GO analyses (Figure 19) categorized

contigs into several biological processes, where catalytic activity is highly represented, indicating an important metabolic activity in *C. consors* venom duct. This may be related to venom production as observed in the case of the scorpion *Parabuthus transvaalicus* where venom replenishing gave rise to about 40% increase in the overall metabolic rate (Nisani et al., 2007). For molecular functions, binding and catalytic activities rank first, which is related to the high conopeptide content of the transcriptome. Finally, three main categories of cellular components were noted: cell, organelle and macromolecular complex, all mostly related to structural proteins involved in secretion and transport of toxic compounds.

3.3. Conopeptides content

Conopeptides were the most abundant compounds of the cone snail venom duct transcriptome with about 42% of the ESTs. In order to retrieve biologically relevant conopeptide sequences, only full length or nearly full length toxins were taken into account in the following work. Sequences with degenerate positions, frameshifts (leading mostly to unusual C-terminal extensions) and sequences without signal peptide were excluded. Finally, the base-called process after sequencing was checked to remove ambiguous positions. This stringent procedure excluded more than 50% of the initial BLASTx-annotated conopeptide candidates. After elimination of redundant sequences, more than 50 conopeptide isoforms were identified. Recent estimations of isoforms number at the proteomic level give a range of 200– 400 (Olivera and Teichert, 2007). Such discrepancy between transcriptomic and proteomic levels in terms of sequence quantity could be explained by post-translational processes (Buczek et al., 2005), but also degradation products, including truncated conopeptides and propeptides resulting from enzymatic cleavage. The stringency of conopeptide annotation procedure used in the transcriptome analysis may also leave aside some sequences.

Some conopeptide sequences were kept confidential for further applications and patents but scaffolds and striking features are provided in the manuscript. Partial data from A superfamily, S superfamily and Conopressins are presented.

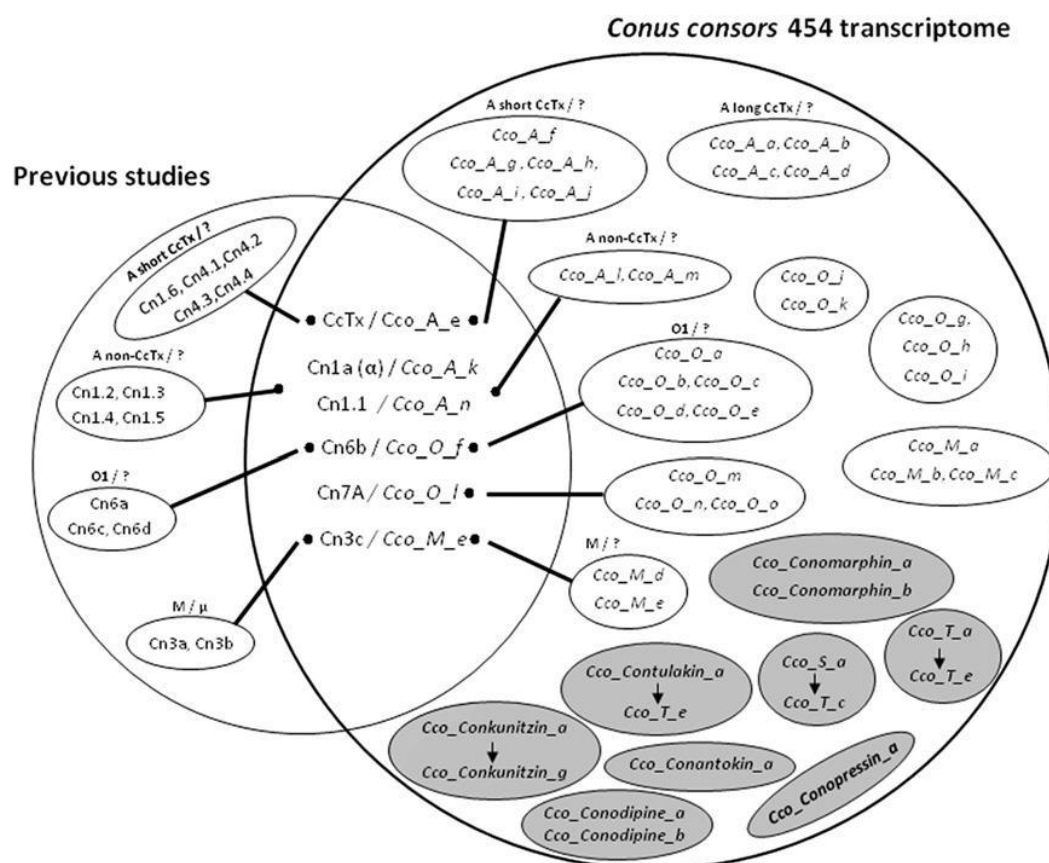


Figure 20: Recovery between 454 transcriptomic identified isoforms and previously identified sequences.

Previous studies include 23 isoforms from three conopeptide families (A, M and O1) (Buczek et al., 2005; Le Gall et al., 1999; Quinton et al., 2009; Liu et al., 2007; Duda and Palumbi, 2004) represented in the left circle. Previous studies include both transcriptomic and proteomic data. Six common isoforms are represented in the recovery area. Related clades including new isoform identified in this study are connected by a line to the common isoform. New families are indicated in grey areas.

Out of the 53 isoforms extracted from this study, only 6 full length precursors have been previously characterized in *C. consors* (Le Gall et al., 1999; Bulaj et al., 2001; Olivera, 2006; Zhang et al., 2006; Biass et al., 2009). Most numerous families found in the transcriptome belong to A, O and M superfamilies. This result is in agreement with recent reports on intraspecific variations observed in cone snail venoms (Jakubowski et al., 2005; Davis et al., 2009; Romeo et al., 2008; Dutertre et al., 2010). These results are summarized in Figure 20.

Table 5: Main features of *Conus consors* conopeptides.

Super-family	Isoform number	Name	Known scaffold(s)	New scaffold(s)	Number of distinct clades	% of the conopeptide cocktail	Remarks
A	14	Cco_A_[a-n]	CC-C-C-C-C (IV) CC-C-C (I)	C-C-C-C- C-CCC- C-C-C-C- C-C	3	82.15%	Mixture of short CcTXs<comma> long CcTXs and canonical isoforms•Primar composant of the transcriptome and composant of the conopeptide cocktail•New scaffold for long CcTXs / potential toxin for short CcTXs
O	15	Cco_O_[a-o]	C-C-CC-C-C (VI/VII)	C-C-CC- C-C-C-C	4	7.57%	4 different clades of O conopeptides with distinct mature sequences•Identification of a new pattern with a potential additional cysteine bound
M	8	Cco_M_[a-h]	Cysteine free (conomarphin) CC-C-C-CC (III)	X	3	6.84%	Presence of an unusual short spacing C4 and (C5-C6)
Conkunitzin	7	Cco_Conkunitzin_[a-g]	C-C-C-C (I)	C-C-C-C- CC-C-C- C-C-C-C	2	1.36%	New pattern related to Kunitz-type proteins involved in toxic function
T	5	Cco_T_[a-e]	Cysteine free	X	2	0.49%	Long and short cysteine-free sequences
Contulakin	4	Cco_Contulakin_[a-d]	NA	X	1	0.79%	Distantely related mature sequence compared

5. Transcriptomics

Super-family	Isoform number	Name	Known scaffold(s)	New scaffold(s)	Number of distinct clades	% of the conopeptide cocktail	Remarks
							Lt1<comma> Lt2 and G contulakins
S	3	Cco_S_[a-c]	C-C-C-C-C- C-C-C-C-C (VIII)	C-C-C-C- C-C-C-C- C-C-C-C	1	0.58%	New isoforms with the most numerous number identified so far among conopeptides
Conodipine	2	Cco_Conodipin_[a-b]	NA	X	1	0.09%	Alpha and beta chains encoded on the mRNA precursor
Conopressin	1	Cco_Conopressin_a	X	C-C-C- CC-C-C	1	0.02%	•First characterization of Pre/Pro conopressin sequence
Conantokin	1	Cco_Conantokin_a	NA	X	1	0.08%	•Presence of 2 cysteines in the C-terminal the sequence (similar to Conantokin-R)
P	1	Cco_P_a	C-C-C-C-C- C (IX)	X	1	0.02%	•Distantly related mature sequence

Detailed results obtained for each family are summarized in Table 5. An important diversity was observed in A superfamily (14 isoforms), O superfamily (15 isoforms), Conkunitzins (7 isoforms) and M superfamily (7 isoforms). Transcripts belonging to the A superfamily were the most abundant sequences found in *C. consors* venom duct transcriptome (Table 5 and Figure 21). 27.5% of the EST reads were identified as part of A conopeptide sequences. Focusing on conopeptides only, members of the A superfamily composed nearly 70% of the transcripts. The A superfamily was divided into two distinct groups based on sequence similarity (CcTx-like conopeptides and a-conotoxins). The first set of sequences included the previously characterized CcTx/Excitotoxin from *C. consors* (Le Gall et al., 1999). Both short (7 isoforms) and long-CcTx (4 isoforms) were detected. Besides the typical CC-C-C-C-C cysteine framework displayed by CcTx, a new cysteine pattern could be observed. Indeed, three isoforms presented an odd number of cysteine (C-C-C-C-C framework)

corresponding to an alteration of the first CC pair. This could potentially lead to the formation of toxins from two monomers as described in a recent publication (Quinton et al., 2009). However, relative transcription level (see further) of these two isoforms remained very low compared to other isoforms (less than 0.01% of the whole conopeptide cocktail). Thus, if biologically relevant, these isoforms probably do not play a key role in the venom function of *C. consors*. The most intriguing feature in the A-superfamily conopeptides was the presence of a long-CcTx isoform with two additional cysteines at the C-terminal part (CC–C–C–C–C–C–C pattern). A high transcription level was observed for this long-CcTx isoform compared to other A-superfamily conopeptides (nearly 18% of the whole conopeptide mixture) confirming the biological relevance of such compounds. Comparison of mature sequences from short- and long-CcTx showed strikingly different primary sequences. Importantly, it was checked that the C-terminal extension did not result from a simple alteration of the stop codon from a short-CcTx. They therefore represent two different clades of toxin as shown by phylogenetic analysis. Besides these original forms of CcTx-related sequences, the more classical α -conotoxin framework CC–C–C represented 4 isoforms. Multiple alignment comparison with known sequences (Figure 22) was used to perform a phylogenetical analysis (Figure 23). This analysis showed that sequences belong to the α -3/5 subfamily, a clade of toxin almost restricted to fish-hunting *Conus* species (Puillandre et al., 2010). One isoform is identical to the mature part of Cn1.1 isolated from *C. consors* which targets muscle nicotinic acetylcholine receptors (Favreau et al., 1999). The three other sequences were closely related to this group with at least 92% identity on the full length sequences.

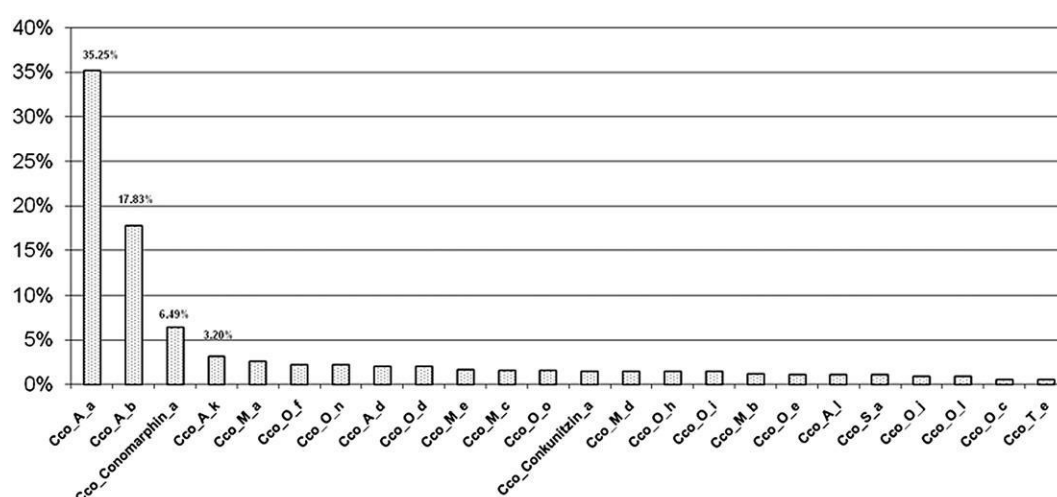


Figure 21: Transcription profiling of the conopeptides cocktail.

Isoforms representing 95% of the conopeptide content are represented on the X axis. Proportion of ESTs (prior to contig assembly step) is represented on the Y axis.

For the O superfamily, a set of 15 isoforms was obtained. Based on sequence similarities and phylogenetic analysis, this set could be divided in 4 different clades. All sequences shared the typical C–C–CC–C–C framework, except one iso-form that presented a C-terminal elongation C–C–CC–C–C– C–C thus giving rise to an original pattern. The relative transcription level analysis of this atypical isoform showed a high expression level in the venom duct. More generally, a great diversity between these different clades that are very different in term of signal and propetide sequences was found (Favreau, 1999; Bulaj et al., 2001). In particular, we identified sequences closely related to LvVIA from the vermivorous *Conus lividus* snail (Duda and Palumbi, 2004), could be interesting candidates for drug design as they clearly separate from other sequences (Favreau, 1999). Previous studies have shown that divergent clades within a toxin-based phylogeny may produce toxins with different functions (Olivera and Teichert, 2007; Conticello et al., 2001).

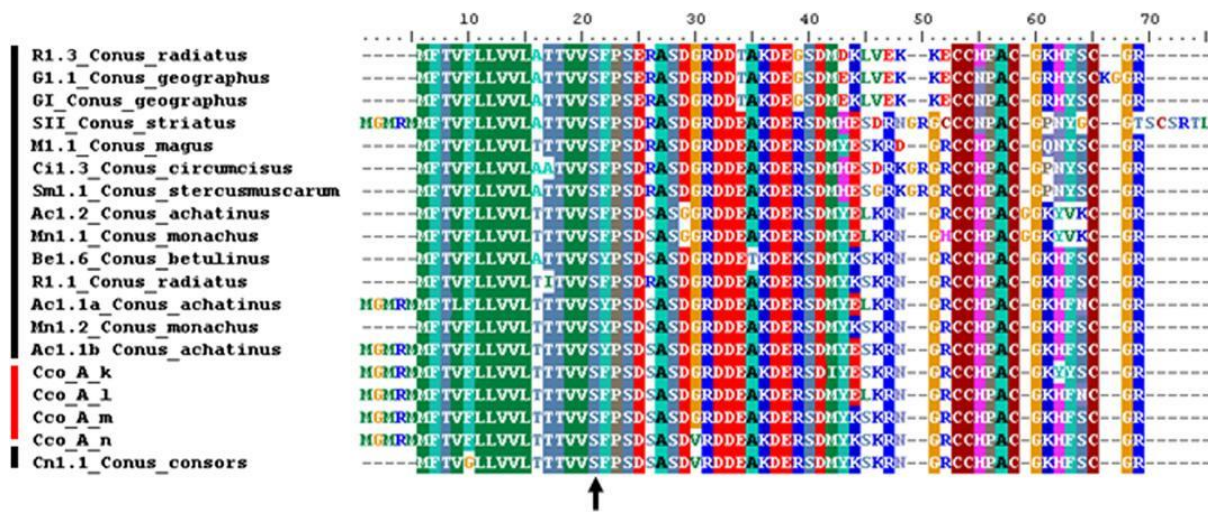


Figure 22: Alignment of A superfamily conopeptides (non CcTx).

New isoforms are indicated by a red line. Cleavage site for the signal peptide is indicated by an arrow. Identity/Similarity for shading is fixed to 80%.

In the M-superfamily, both cysteine-free (2 isoforms) and cysteine-rich (5 isoforms) sequences were detected. The cysteine-free components of this family have been previously named conomorphins (Han et al., 2008) with only 5 sequences described to date in mollusk-hunting species. Indeed, by performing phylogenetical analysis with this group of sequences, all of them were found out to form a monophyletic clade clustering with Marmophin from *Conus marmoreus* but showing significant differences at the protein level. Concerning the cysteine-rich group of conopeptides, the dataset was split into two clades, both belonging to the m-conotoxin family. The first one was related to BullIA, B and C sequences from *Conus bullatus* and the other one included sequences with a shorter spacing between the fourth and fifth cysteine. This last feature could potentially lead to an original three-dimensional structure and subsequently original biological properties.

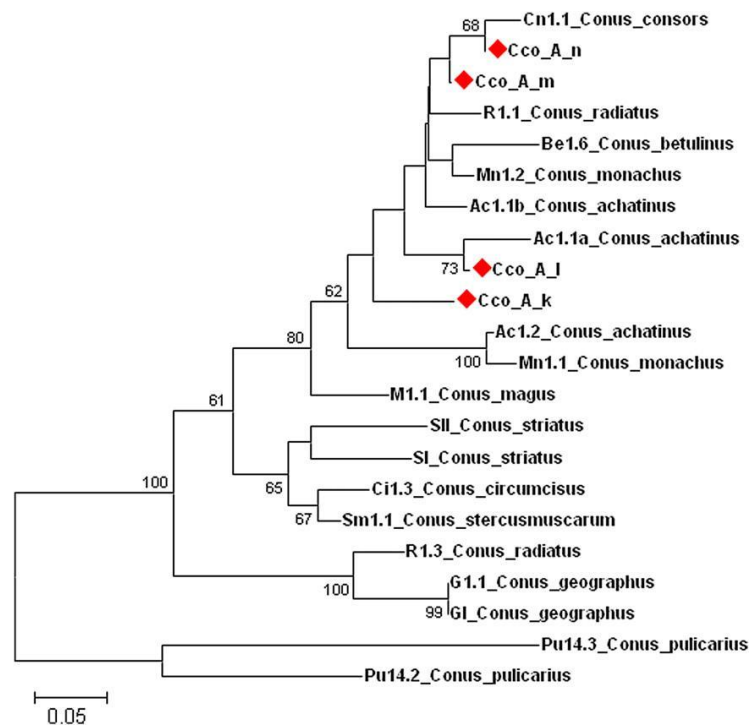


Figure 23: Phylogenetical analysis of *C. consors* A superfamily non-CcTx isoforms.

Dataset includes closely related sequences established from a first large phylo-genetical analysis including all A conopeptides previously characterized. A conopeptides isolated from *Conus pulicarius* have been used as outgroup. Phylogenetical analysis was performed by using the Maximum Likelihood method based on the JTT matrix-based model. The bootstrap consensus tree was inferred from 100 replicates. Bootstrap values above 60% are omitted. A discrete Gamma distribution was used to model evolutionary rate differences among sites (3 categories, a parameter $\frac{1}{4}$ 1.4245). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis includes 64 positions in the final dataset. Sequences retrieved from the current dataset are indicated by a square.

Besides these major superfamilies of conopeptides, many less represented superfamilies and families could also be observed. For instance, 3 konkunitzin isoforms were identified (Bayrhuber et al., 2005; Dy et al., 2006), with a typical cysteine pattern (C–C–C–C). For the first time, 3 konkunitzin-like isoforms were discovered with two original cysteine patterns. The first pattern included an extra cysteine at the C-terminal part. The second one includes two additional cysteine residues (2 sequences) which is a common scaffold for kunitz-type sequences and was also found for the Cal9.1a from *Conus californicus*. A partial precursor

for the conopressin family could be obtained, that provided the first characterization of the related signal sequence. By comparison with known conopressins (Figure 24), the putative mature part of the sequence appeared much longer and displayed at least four additional cysteines (partial sequence) which give a global pattern of C–C–C–CC–C not yet described. However, it cannot be excluded that the maturation process of this precursor would exclude the C-terminal part to produce a more canonical form of conopressin. An alternative hypothesis is to consider the identified conopressin part of this toxin as a byproduct of C-terminal toxin compound. To test it, further studies would have to be performed to show if the small C-terminal product presents a toxic activity on itself or not.

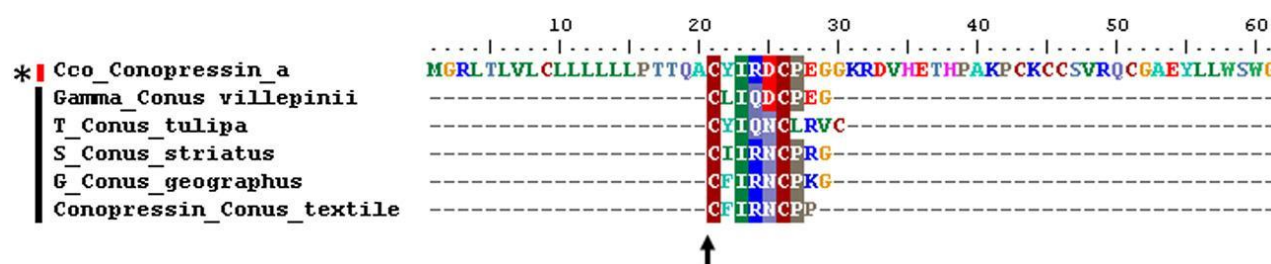


Figure 24: Alignment of S superfamily toxins.

New isoforms isolated from *C. consors* venom ducts are indicated by a red line. New scaffolds are indicated by a star. Cleavage site for the signal peptide is indicated by an arrow. Identity/Similarity threshold for shading is 70%.

For the contulakin family, two closely related isoforms were retrieved from the dataset. These isoforms shared similarities with contulakin–G from *Conus geographus* (Craig et al., 1999) and with contulakin Lt1 and Lt2 from *Conus litteratus* (Pi et al., 2006a).

Concerning the T-superfamily, at least 5 cysteine-free isoforms were found related to this group. With only 12 sequences described to date, this result greatly increases the number of conopeptides from this family.

Two isoforms of conodipine, related to the phospholipase A2 (PLA2) family, have been elucidated and clearly established that alpha and beta chains of the putative full-length sequence are encoded by a unique precursor. To date, only one conodipine sequence has been characterized at the protein level (McIntosh et al., 1995).

A few isoforms of the S-superfamily (Figure 25) and sequences related to the P-superfamily were also observed (Lirazan et al., 2000), as well as some sequences comprising 10 and 12 cysteines that may be related to currently undefined superfamilies (Kaas et al., 2010).

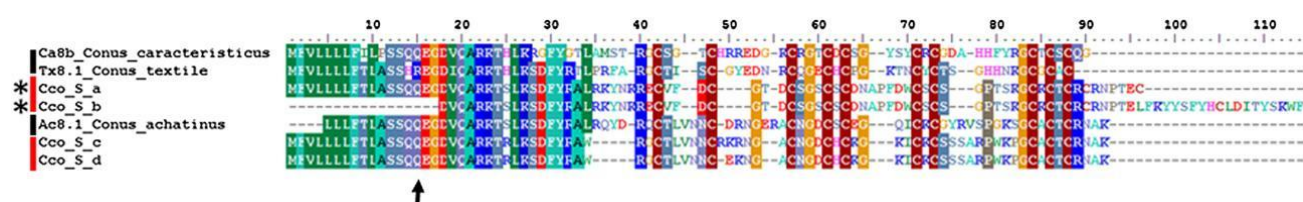


Figure 25: Multiple alignment of Conopressin toxins.

New isoforms isolated from *C. consors* venom ducts are indicated by a red line. New scaffolds are indicated by a star. Cleavage site for the signal peptide is indicated by an arrow. Identity/Similarity threshold for shading is 80%.

Finally, no transcripts deriving from the D, J, L V and Y -superfamilies could be found in this study. This was not surprising as these families have been isolated from vermivorous cone snails only (Kaas et al., 2010). No evidences of I superfamily conopeptides expression was found, which are widespread among piscivorous cone snails.

3.4. Changes in transcript abundance

Huge variations in the conopeptide content were detected which could give new insights on prey capture strategy. Using simple MegaBLAST procedure and appropriate parsing procedure for BLAST hits, major compounds of the conopeptide content were identified. Both major differences in transcription level between different conopeptide superfamilies and

major differences between highly similar isoforms of a given family are highlighted in Figure 21. The conopeptide transcriptional cocktail was composed at one third by the long-CcTx isoform-a. The second top-ranked conopeptides were the long-CcTx isoform-b (new scaffold with two additional cysteines). These two conopeptides represented half of the transcriptional activity of conopeptides. Next to these two main isoforms, the conomorphin isoform-a (no cysteine bonds) represented a relatively small but significant fraction of the conopeptide transcripts (6.49%) compared to the two long-CcTx isoforms. The remaining isoforms had an individual relative transcription level below 5%. Focusing now on the different superfamilies, A superfamily was the main family of the conopeptide cocktail (nearly 70%) and nearly 90% of this cocktail was composed by two long-CcTx isoforms. An opposite situation was observed for the O superfamily conopeptides. Indeed, none of these isoforms belonged to the top-ranked conopeptide transcripts (relative transcriptional level above 5%) but each of the 15 isoforms identified so far represented a similar fraction of the conopeptide cocktail. A similar result was also found for the M and Conkunitzin superfamily conopeptides. Finally, only scarce compounds of the transcriptional content of the venom were representing the T, S, Contulakin and P superfamilies. However, these huge variations between conopeptide expressions at the transcriptional level should be under-stated. Comparison of absolute mRNA transcript abundances with the corresponding protein abundances expressed showed that in many cases mRNA abundance was not a reliable indicator of corresponding protein abundance (Shu and Hong-lui, 2004). Furthermore, attenuation of protein abundance due to post-transcriptional control of protein translations and protein modifications could not be predicted currently from measurement of mRNA abundance

(Wu et al., 2008). Finally, it is noteworthy that the transcriptomic data were produced from three different venom ducts and that consequently the final picture of the venom ducts transcriptome could be the result of slightly different transcript contents for each specimen (Dutertre et al., 2010).

3.5. Toxin-related genes

As mentioned in a recent publication, numerous proteins have been convergently recruited in the venoms of a wide range of taxa (Fry et al., 2009). In the case of *C. consors*, a large array of toxin-related proteins has been unveiled besides conopeptides (Table 6).

Interestingly, actinoporin-like sequences that presented a significant transcription level (3.90%, see Table 6) were discovered. Such sequences have been found in the marine gastropod *Monoplex echo* and named Echotoxin (Shiomi et al., 2002). They may have the ability to bind gangliosides in the erythrocyte membrane and induce subsequent lysis of these cells. They were only distantly related to actinoporin sequences found in sea anemones (Hinds et al., 2002).

One of the most important groups of translated sequences was also composed of hyaluronidases. Such proteins are found in snakes, lizards, stonefish and hymenopterans. They act primarily as diffusion factor, enhancing tissue permeability and consequently the diffusion of the injected venom (Tu and Hendon, 1983). Hyaluronidases identified in *C. consors* transcriptome fell in the same length range of most hyaluronidases with about 450 aa residues. Hyaluronidases present various cysteine patterns (Girish et al., 2002). Two scaffolds were identified in *C. consors* transcriptome (10 and 11 cysteine residues). Sequence comparison of the N-terminal part with previously identified hyaluronidases showed no similarities for the signal peptide domain and low similarities on the C-terminal part encompassing the mature sequence. Consequently, hyaluronidases from *C. consors* formed a distinct clade in the phylogenetic analysis. Hyaluronidases transcripts represented nearly 3.5% of the transcripts and may thus constitute important compounds of the venom cocktail of *C. consors*. Next to hyaluronidases, compounds commonly found in various animal venoms (Fry et al., 2009) such as serine proteases, serine protease inhibitors, metalloproteases and C-type lectins could be recovered from the transcriptome. As most of these compounds are involved in antihemostatic or hemorrhagic activities, they would

Table 6: Main features of toxin-related transcripts.

Name	Status	Length (aa)	Isoform number	Potential effect	% of the raw data hits	Remark
Echotoxin	Partial sequences	NA	3	Pore forming enzymes	3.90%	First descriptionDistantly related to Monoplex echo Echotoxins
Hyaluronidase	Full length and nearly full length	450	4	Venom diffusion factor through enhancement of tissue permeability	3.48%	First descriptionlow similarities compared to known hyaluronidases
Serine protease Serine protease inhibitor	Partial sequence	NA	1	Antihemostatic	0.03%	First decription18 cysteines8 cysteine bonds
Metalloprotease	Full length sequence	307	3	Antihemostatic	0.12%	First description4 cysteines
Defensin	Full lentgh sequence	226	1	Innate immunitu-related antimicrobial peptides	0.01%	First descriptionone additional cysteine bond
C-type lectin	Partial sequence	NA	1	Antihemostatic/myotoxic	0.02%	First descriptionclosely related to the incularin C-type Lectin (land slung)
Cone snail endoprotease (CRIPs)	Full length sequence	NA	4	Toxin and/or proteolytic processing of venom compounds	0.07%	First description45 cysteines residuesdistantly related to Mr30 and Tex31
Ferritin	Full length sequence	NA	≥10	Unknown	21.50%	Highly similar to Aplysia californica and Lottia gigantea proteins
Disulfide isomerase	Full length sequence	357	2	Folding of cysteine-rich proteins	0.17%	Highly similar to th Conus marmoreus disulfide isomerase

complete the diffusion enhancement of conopeptides besides hyaluronidases. Sequences related to previously characterized cone snail Mr30 and Tex31 endoproteases (Qian et al., 2008) were identified. Consensus sequence from *C. consors* presented 24 additional cysteines compared to Tex31 and MR30 for a total of 45 cysteines on nearly 500 aa in length. This sequence constitutes a really original sequence and the real biological function remains to be clarified especially in the frame of proteolytic activity responsible for processing of pro-conopeptides. Finally, two isoforms similar to defensin-related proteins have been identified. Most of defensin-like proteins present antimicrobial activities and are commonly found in snakes, scorpions (Gao et al., 2010) and spiders venoms (Yan and Adams, 1998) as well as bivalve mollusks (Zhao et al., 2007). The actual function of anti-microbial peptides in these venoms is currently not well understood. One hypothesis is that they may protect predatory animals from microorganism infections by keeping their prey sterile (Zhao et al., 2011). Thus, further work would be necessary to identify the actual function of these proteins. The major features of these new families of compounds described from a cone snail venom duct transcriptome are summarized in Table 6.

Besides these proteins which may have toxic activities themselves, several other protein families have been identified that could be involved in the maturation processes of conopeptides, such as disulfide isomerases (Zhao et al., 2007). These proteins are of great interest as they could help to better understand the conopeptide maturation and folding processes, information that would be useful for chemical synthesis.

One intriguing feature of the *C. consors* venom duct transcriptome was the presence of more than 20% of ferritin-related transcripts. No signal peptides were identified but these proteins should play a key role in the cellular production of the venom compounds. To date, ferritin proteins are not specifically mentioned in venom apparatus, neither at the transcriptomic nor proteomic level. A hypothesis is that such high ferritins level may sequester free iron resulting from high metabolic activity of venom duct cells.

3.6. Xenobiotics

Pyrosequencing from metazoan tissues can produce sequences from symbionts (Vera et al., 2008) also called xenobiotics. Of the cone snail venom duct trimmed ESTs, sequences associated with a closest BLAST hit from various clades of arche or eubacteria were detected. Major families of bacterial sequences belonged to firmicutes clade (54 sequences, most of them belonging to streptococcus subgroup), gamma proteobacteria (18 sequences), alpha proteobacteria (16 sequences) and beta proteobacteria (8 sequences) based on taxonomy classification of the NCBI database. Three sequences clearly belonged to the actinomycetes group of bacteria. One sequence from an archaea bacterium was also identified. Most of these sequences are related to metabolism or classical cellular functions and no evidences for potential implication in the venom function could be inferred. External contaminations of the cone venom duct samples cannot be excluded but recent microbiological studies detected numerous associated bacteria and microbial communities in cone snails, especially actinomycetes (Peraud et al., 2009; Lin et al., 2010) and comparative results have also been obtained in closely related taxa (Galindo et al., 2010).

3.7. Transposable elements

Transposable elements (TE) occupy a large fraction of many multicellular eukaryotic genomes, ranging from 4% in the yeast *Saccharomyces cerevisiae*, to 45% in humans (Biémont and Vieira, 2006) and more than 70% in some plants and amphibians. Their movement and accumulation represent a major force shaping the genes and genomes of almost all organisms.

It is generally accepted that the wide variation in genome size observed among eukaryotic species is more closely correlated with the amount of repetitive DNA than with the number of coding genes (Kidwell, 2002). As genome size estimation for Neogastropoda is quite similar

to human genome size (3.1 pg/haploid genome estimated on 26 species, standard error 0.87; 3.5 pg/haploid genome for human, Gregory et al., 2007), nearly half of the *C. consors* genome should be occupied by TEs as in human genome. Despite a low fraction of contigs identified as TEs (1.6%), major families of TEs were found, including class I TEs (which act through an RNA intermediate) from Harbinger, TC1/mariner, piggyback and Polinton families. Class II TEs (which also act through a DNA intermediate) included both non-LTR retrotransposons from various clades of LINE (L1, RTE, I, L2, RTE) and LTR-retrotransposons from Ty3/Gypsy, Ty1/Copia, and DIRS families. While no evidences were found for preferential association of TEs with venom compounds of the *C. consors* cone snail, undergoing full genome sequencing project will certainly provide new insights for the potential role of TEs in the venom production.

By performing phylogenetic analysis on part of TEs, a striking feature could be evidenced. Most of the sequences appear to cluster with distantly related taxonomic groups living in fresh water or marine environments. This result had been previously been highlighted for other marine species (Terrat et al., 2008) and could be the consequence of frequent horizontal transfers of TEs among species living in the same environment, probably due to parasite/host interactions (Gilbert et al., 2010).

4. Discussion and conclusion

4.1. Discussion

The results presented in this study offer an unprecedented amount of information on the venom duct transcriptome of a cone snail. The sensitivity and unbiased nature of the sample preparation gave access to both the main toxin-related compounds and to more rare transcripts. In a relatively short timeframe, many new conopeptides have been discovered with at least 53 isoforms in various superfamilies and families. Study of transcripts

abundance showed that nearly half of the venom duct transcriptome was composed of two long-CcTx isoforms and hyaluronidase proteins. Nonetheless, numerous families of compounds were present at a medium to low-level in the transcriptome. Whether or not they play a role in the venom function in these specimens is a question to address. They may also constitute a stock of optional weapons which could be quickly produced to capture different preys.

Whether these additional compounds of the venom are biologically active is a question to address. Comparison of the transcriptomic and proteomic could help to validate new toxin candidates. In addition to this transcriptomic study, an in-depth venom proteomic work has been performed using various techniques, that will eventually provide a global view of the venom content in terms of protein sequences and families (manuscript in preparation).

This study also gave valuable information for unusual or potentially novel toxins and their related isoforms that were difficult to detect in the proteome (Harrison et al., 2007). One interesting point with the 454 pyrosequencing method is the possibility to detect very low abundance transcripts which would be certainly missed with a classical cDNA cloning approach (Pi et al., 2006a, 2006b; Biggs et al., 2010) even using a non-normalized library. This study also demonstrated that the convergent recruitment of numerous protein families for venom purpose is also applying to cone snails (Fry et al., 2009). Beyond conopeptides, the presence of transposable elements and sequences from distantly related species raised the question of their potential role in the venom function. TEs have been previously identified in PLA2 genes from the venom gland of *Vipera ammodytes* (Kordis and Gubensek, 1998). Most transposition events involving protein-coding regions are deleterious to host genomes. However since TEs associated with PLA2 genes have been found to be transcribed, the TEs in *C. consors* may hypothetically act as a source of genetic novelty by playing an important role in the origin and diversity of toxin genes.

4.2. Conclusions

This study remains a picture based on three cone snail venom ducts at a given time. A recent study using mass spectrometry based on the same species showed that dramatic intraspecimen variations at the proteomic level do exist (Dutertre et al., 2010). A single cone snail could produce two very distinct venom profiles containing completely different sets of peptides with no overlap of detected masses at two different times. In this case, one might expect a very different transcriptomic picture than that obtained in this study. The huge diversity of toxin compounds at the level of a single species will be addressed by comparing genomic, transcriptomic and proteomic data. This may allow identifying the mechanisms involved in selection and regulation of the venom compounds and the importance of transcriptional and post-transcriptional regulation levels. Such studies with various cone snails will allow unraveling complex mechanisms which lead to the remarkable variability of the venom composition observed within the cone snail group.

Acknowledgment

This work has been supported by a grant of the Euro-pean Commission: CONCO, the cone snail genome project for health. Integrated Project ref. LSHB-CT-2007-037592 (<http://www.conco.eu>).

Conflict of interest statement

The authors declare that there are no conflicts of interest.

References

- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105.
- Altschul, S.F., et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Ashburner, M., et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25, 25–29.
- Bayrhuber, M., et al., 2005. Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. *The Journal of Biological Chemistry* 280, 23766–23770.
- Blass, D., et al., 2009. Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *Journal of Proteomics* 72, 210–218.
- Biémont, C., Vieira, C., 2006. Genetics: junk DNA as an evolutionary force. *Nature* 443, 521–524.
- Biggs, J.S., et al., 2010. Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Molecular Phylogenetics and Evolution* 56, 1–12.
- Buczek, O., Bulaj, G., Olivera, B.M., 2005. Conotoxins and the post-translational modification of secreted gene products. *Cellular and Molecular Life Sciences* 62, 3067–3079.
- Bulaj, G., et al., 2001. Delta-conotoxin structure/function through a cladistic analysis. *Biochemistry* 40, 13201–13208.
- Conesa, A., Götz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008, 619832.

- Conticello, S.G., et al., 2001. Mechanisms for evolving hyper-variability: the case of conopeptides. *Molecular Biology and Evolution* 18, 120–131.
- Craig, A.G., et al., 1999. Contulakin-G, an O-glycosylated invertebrate neurotensin. *The Journal of Biological Chemistry* 274, 13752–13759.
- Craig, A.G., et al., 2001. Enzymatic glycosylation of contulakin-G, a glyco-peptide isolated from *Conus* venom, with a mammalian ppGalNAc-transferase. *Toxicon* 39, 809–815.
- Davis, J., Jones, A., Lewis, R.J., 2009. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 30, 1222–1227.
- Duda, T.F., Palumbi, S.R., 2004. Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. *Proceedings. Biological Sciences/The Royal Society* 271, 1165–1174.
- Dutertre, S., et al., 2010. Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* 55, 1453–1462.
- Dy, C.Y., et al., 2006. Structure of conkunitzin-S1, a neurotoxin and Kunitz-fold disulfide variant from cone snail. *Acta Crystallographica. Section D, Biological Crystallography* 62, 980–990.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Favreau, P., et al., 1999. Biochemical characterization and nuclear magnetic resonance structure of novel alpha-conotoxins isolated from the venom of *Conus consors*. *Biochemistry* 38, 6317–6326.
- Favreau, P., 1999. Physico-chemical and pharmacological study of new toxins from the piscivorous cone snail *Conus consors*. PhD thesis.

- Favreau, P., Stöcklin, R., 2009. Marine snail venoms: use and trends in receptor and channel neuropharmacology. *Current Opinion in Pharmacology* 9, 594–601.
- Fry, B.G., et al., 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annual Review of Genomics and Human Genetics* 10, 483–511.
- Galindo, J., Grahame, J.W., Butlin, R.K., 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology* 23, 2004–2016.
- Gao, B., et al., 2010. A potent potassium channel blocker from *Mesobuthus eupeus* scorpion venom. *Biochimie* 92, 1847–1853.
- Gilbert, C., et al., 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464, 1347–1350.
- Girish, K.S., et al., 2002. Snake venom hyaluronidase: an evidence for isoforms and extracellular matrix degradation. *Molecular and Cellular Biochemistry* 240, 105–110.
- Gregory, T.R., et al., 2007. Eukaryotic genome size databases. *Nucleic Acid Research* 35, 332–338.
- Guindon, S., et al., 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology* 537, 113–137.
- Harrison, R.A., et al., 2007. Identification of cDNAs encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and unusually short variant transcripts. *Gene* 392, 22–33.
- Han, Y., et al., 2008. Purification and structural characterization of a D-amino acid-containing conopeptide, conomarphin, from *Conus marmoreus*. *The FEBS Journal* 275, 1976–1987.

- Hinds, M.G., et al., 2002. Solution structure of the eukaryotic pore-forming cytolytic equinatoxin II: implications for pore formation. *Journal of Molecular Biology* 315, 1219–1229.
- Huang, Y., et al., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682.
- Jakubowski, J.A., et al., 2005. Intraspecific variation of venom injected by fish-hunting *Conus* snails. *The Journal of Experimental Biology* 208, 2873–2883.
- Jurka, J., et al., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110, 462–467.
- Kaas, Q., Westermann, J.-C., Craik, D.J., 2010. Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon* 55, 1491–1509.
- Kauferstein, S., et al., 2011. Venomic study on cone snails from South Africa. *Toxicon* 57, 28–34.
- Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115, 49–63.
- Kordis, D., Gubensek, F., 1998. The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes. *Toxicon* 36, 1585–1590.
- Le Gall, F., et al., 1999. A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na^+ -dependent mechanism. *The European Journal of Neuroscience* 11, 3134–3142.
- Lin, Z., et al., 2010. Pulicatins A–E, Neuroactive Thiazoline Metabolites from cone snail-associated bacteria. *Journal of Natural Products* 73, 1922–1926.
- Lirazan, M.B., et al., 2000. The spasmodic peptide defines a new conotoxin superfamily. *Biochemistry* 39, 1583–1588.

- Liu, L., et al., 2007. Two potent alpha3/5 conotoxins from piscivorous *Conus achatinus*. *Acta Biochimica Et Biophysica Sinica* 39, 438–444.
- Lopez-Vera, E., et al., 2008. Role of hydroxyprolines in the in vitro oxidative folding and biological activity of conotoxins. *Biochemistry* 47, 1741–1751.
- McIntosh, J.M., et al., 1995. Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*. *The Journal of Biological Chemistry* 270, 3518–3526.
- Nisani, Z., Dunbar, S.G., Hayes, W.K., 2007. Cost of venom regeneration in *Parabuthus transvaalicus*. *Comparative Biochemistry and Physiology. Part A. Molecular & Integrative Physiology* 147, 509–513.
- Olivera, B.M., 2006. *Conus* peptides: biodiversity-based discovery and exogenomics. *The Journal of Biological Chemistry* 281, 31173–31177.
- Olivera, B.M., Teichert, R.W., 2007. Diversity of the neurotoxic *Conus* peptides: a model for concerted pharmacological discovery. *Molecular Interventions* 7, 251–260.
- Parameswaran, P., et al., 2007. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multi-plexing. *Nucleic Acids Research* 35, e130.
- Parkinson, J., Blaxter, M., 2004. Expressed sequence tags: analysis and annotation. *Methods in Molecular Biology* 270, 93–126.
- Peraud, O., et al., 2009. Microhabitats within venomous cone snails contain diverse actinobacteria. *Applied and Environmental Microbiology* 75, 6820–6826.
- Pi, C., Liu, J., et al., 2006a. Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics* 88, 809–819.

- Pi, C., Liu, Y., et al., 2006b. Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie* 88, 131–140.
- Pisarewicz, K., et al., 2005. Polypeptide chains containing D-gamma-hydroxyvaline. *Journal of the American Chemical Society* 127, 6207– 6215.
- Posada, D., 2009. Selection of models of DNA evolution with jModelTest. *Methods in Molecular Biology* 537, 93–112.
- Puillandre, N., Watkins, M., Olivera, B.M., 2010. Evolution of *Conus* peptide genes: Duplication and Positive selection in the A-Super-family. *Journal of Molecular Evolution* 70, 190–202.
- Qian, J., Guo, Z.-yun, Chi, C.-wu, 2008. Cloning and isolation of a *Conus* cysteine-rich protein homologous to Tex31 but without proteolytic activity. *Acta Biochimica Et Biophysica Sinica* 40, 174–181.
- Quinton, L., Gilles, N., De Pauw, E., 2009. TxXIIIA, an atypical homodimeric conotoxin found in the *Conus textile* venom. *Journal of Proteomics* 72, 219–226.
- Romeo, C., et al., 2008. *Conus ventricosus* venom peptides profiling by HPLC-MS: a new insight in the intraspecific variation. *Journal of Separation Science* 31, 488–498
- Shiomi, K., et al., 2002. Properties of proteinaceous toxins in the salivary gland of the marine gastropod. *Toxicon* 40, 563–571.
- Shu, Y., Hong-Hui, L., 2004. Transcription, translation, degradation, and circadian clock. *Biochemical and Biophysical Research Communications* 321, 1–6.
- Stanhope, M.J., et al., 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411, 940– 944.

- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56, 564–577.
- Tamura, K., et al., 2007. MEGA4: molecular evolutionary genetics analysis software version 4.0. *Molecular Biology and Evolution* 24, 1596–1599.
- Terlau, H., Olivera, B.M., 2004. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiological Reviews* 84, 41–68.
- Terrat, Y., Bonnivard, E., Higuete, D., 2008. GalEa retrotransposons from galatheid squat lobsters define a new clade of Ty1/copia-like elements restricted to aquatic species. *Molecular Genetics and Genomics: MGG* 279, 63–73.
- Tu, A.T., Hendon, R.R., 1983. Characterization of lizard venom hyaluronidase and evidence for its action as a spreading factor. *Comparative Biochemistry and Physiology. B, Comparative Biochemistry* 76, 377–383.
- UniProt Consortium, 2011. Ongoing and future developments at the Universal protein Resource. *Nucleic Acids Research* 39 (Database issue), D214–D219.
- Vera, J.C., et al., 2008. Rapid transcriptome characterization for a non-model organism using 454 pyrosequencing. *Molecular Ecology* 17, 1636–1647.
- Waterhouse, A.M., et al., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Wu, G., Nie, L., Zhang, Weiwen, 2008. Integrative analyses of post-transcriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Current Microbiology* 57, 18–22.
- Yan, L., Adams, M.E., 1998. Lycotoxins, antimicrobial peptides from venom of the wolf spider *Lycosa carolinensis*. *The Journal of Biological Chemistry* 273, 2059–2066.

- Zhang, M.-M., et al., 2006. Structural and functional diversities among mu-conotoxins targeting TTX-resistant sodium channels. *Biochemistry* 45, 3723–3732.
- Zhang, Z., et al., 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 7, 203–214.
- Zhao, H., et al., 2011. A defensin-like antimicrobial peptide from the venoms of spider, *Ornithoctonus hainana*. *Journal of Peptide Science*.
- Zhao, J., et al., 2007. Molecular cloning, expression of a big defensin gene from bay scallop *Argopecten irradians* and the antimicrobial activity of its recombinant protein. *Molecular Immunology* 44, 360–368.

5.4 Concluding remarks

The high-quality preparation of the *C. consors* venom duct samples and the second-generation sequencing technique used for this work both gave access to a phenomenal amount of data. As for conopeptides, this enabled a total of 53 distinct isoforms to be described in the different superfamilies. Two major components, in terms of transcripts abundance, seemed to stand out: CcTx toxins on one hand and hyaluronidases on the other – also further confirmed by their high presence in the venom proteome. The CcTx toxin had already been the subject of a specific analysis¹¹, as did the hyaluronidases after this work¹².

The sequencing method also enables the transcripts that were expressed in very low quantities to be detected – which would have not been possible with cDNA cloning techniques – hence providing a much better sequence coverage.

Of course, the other main known conopeptide superfamilies and families and other peptide and protein families were also represented, providing a very detailed overview of all the components expressed in the *C. consors* venom duct transcriptome.

Thus, many new cone snail transcriptome analyses were undertaken¹³⁻¹⁷, proving that the concept is quite innovative and very important for the rapid discovery of new component families and therefore of future potential drug candidates.

At this stage, the comparison with the venom proteome and peptidome will be necessary to address the questions of whether the additional components found in the transcriptome, next to the common toxins usually found in cone snail venoms, are biologically active and of any use in the venomous function.

5.5 Reference List

1. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database: its relevance to human molecular medical research. *J. Mol. Med. (Berl)* 1997, 75 (5), 312-316.
2. Sanger, F.; Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 1975, 94 (3), 441-448.
3. Hillyard, D. R.; Olivera, B. M.; Woodward, S.; Corpuz, G. P.; Gray, W. R.; Ramilo, C. A.; Cruz, L. J. A molluscivorous *Conus* toxin: conserved frameworks in conotoxins. *Biochemistry* 1989, 28 (1), 358-361.
4. Woodward, S. R.; Cruz, L. J.; Olivera, B. M.; Hillyard, D. R. Constant and hypervariable regions in conotoxin propeptides. *EMBO J.* 1990, 9 (4), 1015-1020.
5. Shon, K. J.; Grilley, M. M.; Marsh, M.; Yoshikami, D.; Hall, A. R.; Kurz, B.; Gray, W. R.; Imperial, J. S.; Hillyard, D. R.; Olivera, B. M. Purification, characterization, synthesis, and cloning of the lockjaw peptide from *Conus purpurascens* venom. *Biochemistry* 1995, 34 (15), 4913-4918.
6. Olivera, B. M.; Walker, C.; Cartier, G. E.; Hooper, D.; Santos, A. D.; Schoenfeld, R.; Shetty, R.; Watkins, M.; Bandyopadhyay, P.; Hillyard, D. R. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann. N. Y. Acad. Sci.* 1999, 870 (MOLECULAR STRATEGIES IN BIOLOGICAL EVOLUTION), 223-237.
7. Pi, C.; Liu, Y.; Peng, C.; Jiang, X.; Liu, J.; Xu, B.; Yu, X.; Yu, Y.; Jiang, X.; Wang, L.; Dong, M.; Chen, S.; Xu, A. L. Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie* 2006, 88 (2), 131-140.
8. Pi, C.; Liu, J.; Peng, C.; Liu, Y.; Jiang, X.; Zhao, Y.; Tang, S.; Wang, L.; Dong, M.; Chen, S.; Xu, A. Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics* 2006, 88 (6), 809-819.

9. Hu, H.; Bandyopadhyay, P. K.; Olivera, B.; Yandell, M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC. Genomics* 2011, *12* (1), 60.
10. Terrat, Y.; Biass, D.; Dutertre, S.; Favreau, P.; Remm, M.; Stocklin, R.; Piquemal, D.; Ducancel, F. High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon* 2012, *59* (1), 34-46.
11. Le, G. F.; Favreau, P.; Benoit, E.; Mattei, C.; Bouet, F.; Menou, J. L.; Menez, A.; Letourneux, Y.; Molgo, J. A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na⁺-dependent mechanism. *Eur. J. Neurosci.* 1999, *11* (9), 3134-3142.
12. Violette, A.; Leonardi, A.; Piquemal, D.; Terrat, Y.; Biass, D.; Dutertre, S.; Noguier, F.; Ducancel, F.; Stocklin, R.; Krizaj, I.; Favreau, P. Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar. Drugs* 2012, *10* (2), 258-280.
13. Lluisma, A. O.; Milash, B. A.; Moore, B.; Olivera, B. M.; Bandyopadhyay, P. K. Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar. Genomics* 2012, *5*, 43-51.
14. Hu, H.; Bandyopadhyay, P. K.; Olivera, B. M.; Yandell, M. Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC. Genomics* 2012, *13* (1), 284.
15. Jin, A. H.; Dutertre, S.; Kaas, Q.; Lavergne, V.; Kubala, P.; Lewis, R. J.; Alewood, P. F. Transcriptomic Messiness in the Venom Duct of *Conus miles* Contributes to Conotoxin Diversity. *Mol. Cell Proteomics.* 2013, *12* (12), 3824-3833.
16. Lavergne, V.; Dutertre, S.; Jin, A. H.; Lewis, R. J.; Taft, R. J.; Alewood, P. F. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC. Genomics* 2013, *14* (1), 708.
17. Safavi-Hemami, H.; Hu, H.; Gorasia, D. G.; Bandyopadhyay, P. K.; Veith, P. D.; Young, N. D.; Reynolds, E. C.; Yandell, M.; Olivera, B. M.; Purcell, A. W. Combined Proteomic and Transcriptomic Interrogation of the Venom Gland of *Conus geographus*

Uncovers Novel Components and Functional Compartmentalization. *Mol. Cell Proteomics*. 2014, 13 (4), 938-953.

6. Peptidomics and Proteomics

6.1 Introduction

As highlighted in the previous chapter, the sheer volume of proteomic and peptidomic data requiring analysis has become problematic. To address this issue and extract the most pertinent information from the wealth of data yielded by proteomes and peptidomes, the use of transcriptomes and bioinformatics has proven to be the most adequate solution in terms of results and time.

6.2 Bioinformatics for transcriptome

In order to be of further assistance in the analysis of proteomic and peptidomic data, the transcriptomic data themselves need to be processed, not only to directly extract sequence information but also to assess the sequencing quality.

Recent advances in molecular biology sequencing technologies provide scientists with very comprehensive transcriptomic data. Unfortunately, manually cleaning and classifying all data could soon become a real hassle. Luckily, bioinformatic tools have come to the rescue in the extraction of relevant information from transcriptomes.

In the case of venom transcriptomes, the first step in the processing of raw data obtained by next-generation sequencing is the *de novo* assembly. Indeed, since very often no genomic data is available for sequence mapping, a ground-up assembly is unavoidable. This process consists of transforming raw sequences into contigs, the main goal being to remove redundancy and build a library of unique sequences. This step is often achieved by

proprietary software packages provided with the sequencing instrument. This was the case with Newbler, the software used for the *Conus consors* transcriptome assembly, and supplied with the Roche 454 pyrosequencer when the instrument was sold at the time. Since then, many other new assembler programmes have been created – proprietary and open source, general or specific to certain applications – to overcome the huge amount of data generated by next-generation sequencing instruments. This step is of course essential since it is the base of any further analysis, either directly in the transcriptome data or in combination with proteomic data. Several studies have been undertaken to provide scientists with a comparison of the most pertinent assembling programmes ^{1,2}.

The next step is the extraction of the information from the transcriptomes. Dozens of tools exist that can be more or less complex depending on the data that needs to be analysed. Until recently, in the field of venomomics (and many others, for that matter), no specific bioinformatic tools were available. Most work had to be undertaken by manually searching databases using classical BLAST searches and GO annotations for example (see Figure 26).

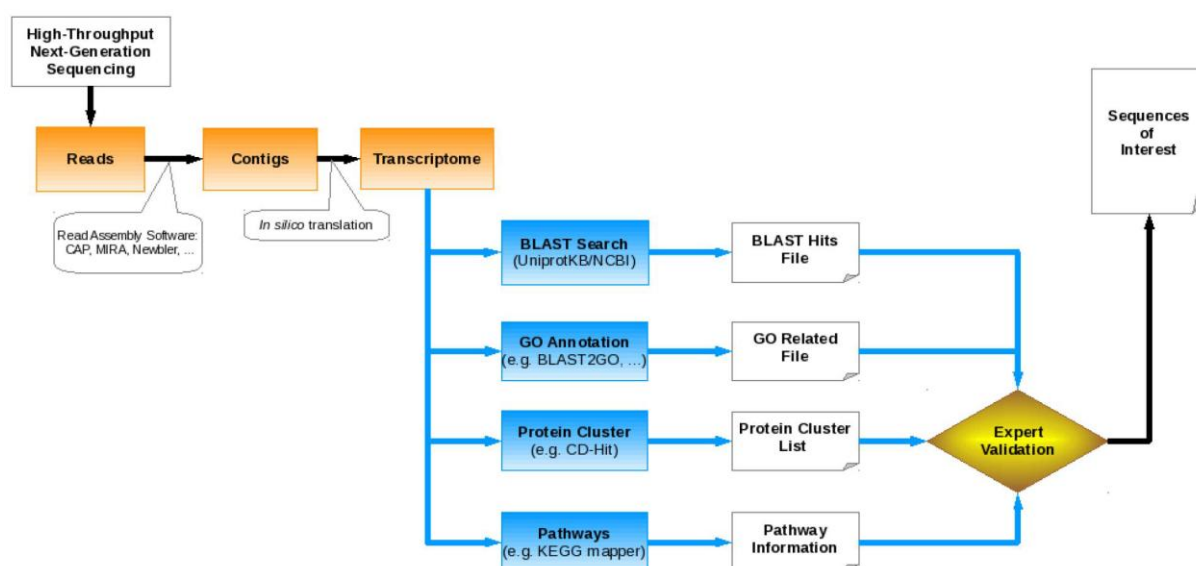


Figure 26: Summary of classical transcriptome analysis operations.

Source: PhD Thesis of KOUA Dominique Kadio: Bioinformatics tools to assist drug candidate discovery in venom gland transcriptomes ³.

In order to circumvent this hindrance, one the methods used in the context of the CONCO project was to establish a bioinformatic platform for the collection, classification and analysis of transcriptomic and proteomic data. Bioinformatic tools and a user-friendly platform dedicated to the analysis of cone snail venom transcriptomes were created with a view to rapidly and semi-automatically obtain valuable sequences information, properly classified by superfamily or family ³⁻⁶. The platform also ensures the integration of proteomic data in an adequate classified manner (see Figure 27). The methods implemented for these tools could to some extent be quite easily transposed to the transcriptomic analysis of other venomous species.

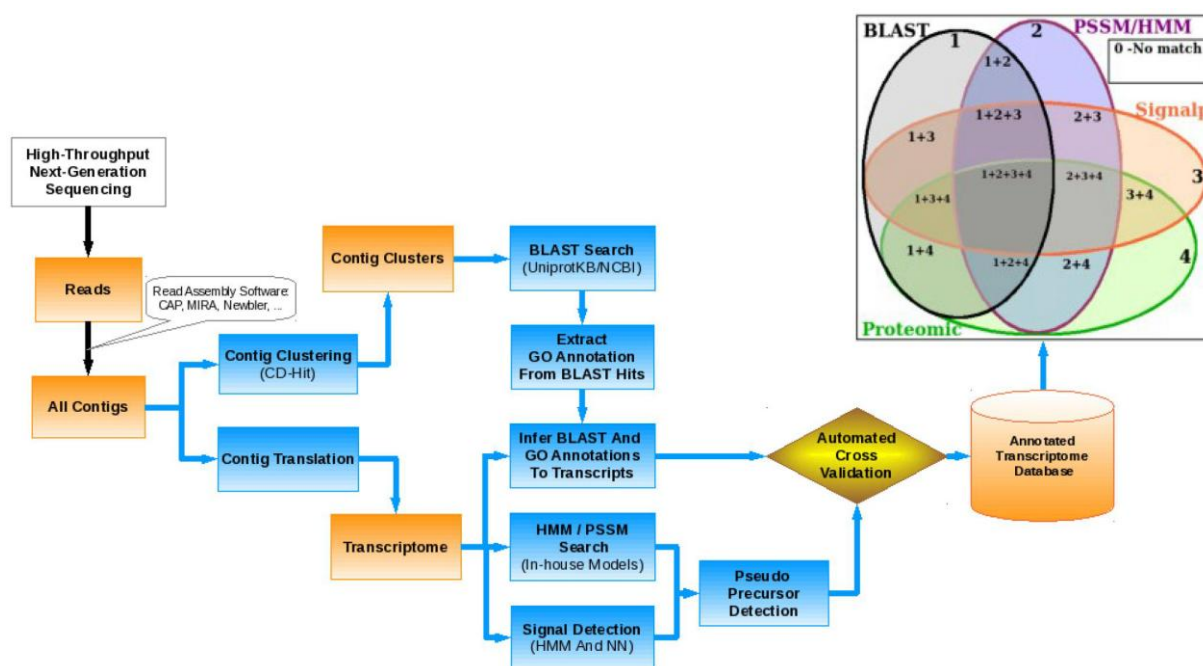


Figure 27: Improved workflow for transcriptome analysis.

Modified from source: PhD Thesis of KOUA Dominique Kadio: Bioinformatics tools to assist drug candidate discovery in venom gland transcriptomes ³.

6.3 Bioinformatics for peptidome and proteome

With transcriptomic data duly analysed, classified and the valuable sequence information obtained, why bother continue work on the peptidomics and proteomics? For the same reasons that biological cellular processes have not stopped to RNA transcripts. Peptides and proteins will undergo further modifications – i.e. post-translational modifications (PTMs) – that are very often essential for their final structure and activity. With the exception of a few predictive tools and homology with known sequences, using the sequence information directly from transcriptomes could be problematic to find correct functional activities for potential drug candidates for example. MS and MS/MS data are therefore essential to generate further information on these PTMs. Detection of those modifications can, however, prove troublesome since cone snail venoms are known to possess many post-translational modified peptides and proteins^{7,8}. Many studies have proved that PTMs are essential for the conotoxins activity and that their selectivity and efficacy is often ruled by these modifications⁹⁻¹¹.

Finding PTMs by *de novo* MS/MS adds to the intrinsic complexity of the actual sequencing and is therefore even more time consuming. Even so, PTMs are quite complex to implement in programmes, thus the use of bioinformatic tools should prove handy to assist the MS/MS sequencing¹².

MS/MS fragmentation fingerprinting tools, like Sequest¹³, X!Tandem¹⁴, Mascot¹⁵ and Phenyx¹⁶, are used to link the proteomic and peptidomic data to different databases. In the case of the three following studies, Phenyx was used to match the transcriptomic data with the peptidomic and proteomic data. To achieve this, the programme randomly splits the transcriptome contigs into small fragments and generates a database of theoretical MS/MS spectra. The latter are then matched with the experimental MS/MS spectra obtained (see

Figure 28), but without forgetting to include the different PTMs that may occur with conotoxins or venom peptides in general.

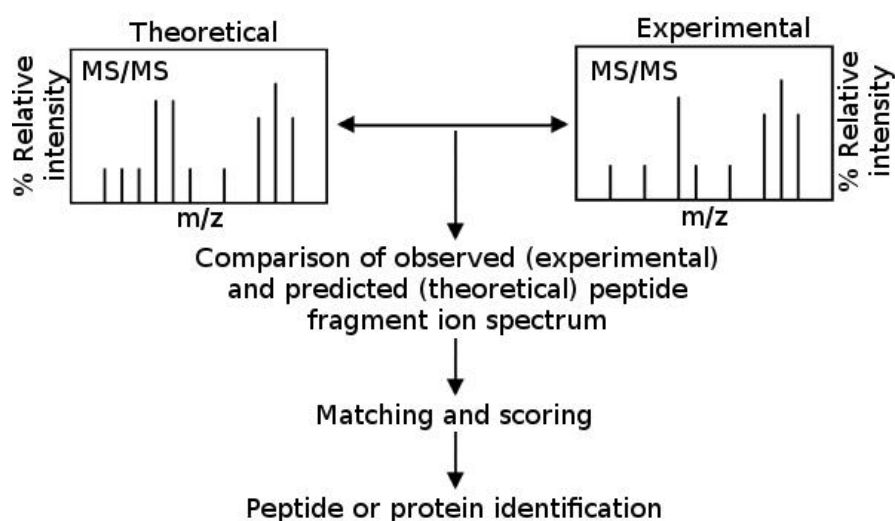


Figure 28: Scheme of the workflow used for the peptide or protein identification.

6.4 Transcriptomics to speed up proteomics and peptidomics

The three following articles describe methods that are based on the same principle: the use of the venom gland transcriptome data to directly assist peptidome and proteome analyses. The main differences between the methods lie in the sample source and preparation in an effort to maximise the comprehensiveness of the venom content map of a cone snail, namely *C. consors*.

Large-scale discovery of conopeptides and conoproteins in the injected venom of a fish-hunting cone snail using a combination of proteomics and transcriptomics.

Aude Violette¹, Daniel Biass¹, Sébastien Dutertre¹, Dominique Koua¹, David Piquemal², Fabien Pierrat², Reto Stöcklin¹, Philippe Favreau¹.

¹ Atheris Laboratories, case postale 314, 1233 Bernex-Geneva, Switzerland

² Skuldtech, 134 rue du Curat, Bat. Amarante, 34090 Montpellier, France

Publication data

Published the 13.06.2012 in the Journal of Proteomics, Vol. 75, Issue 17, p.5215-5225.

Pubmed ID: 22705119

DOI: 10.1016/j.jprot.2012.06.001

Cited 31 times on Google Scholar, 26 times on ScienceDirect.

Keywords

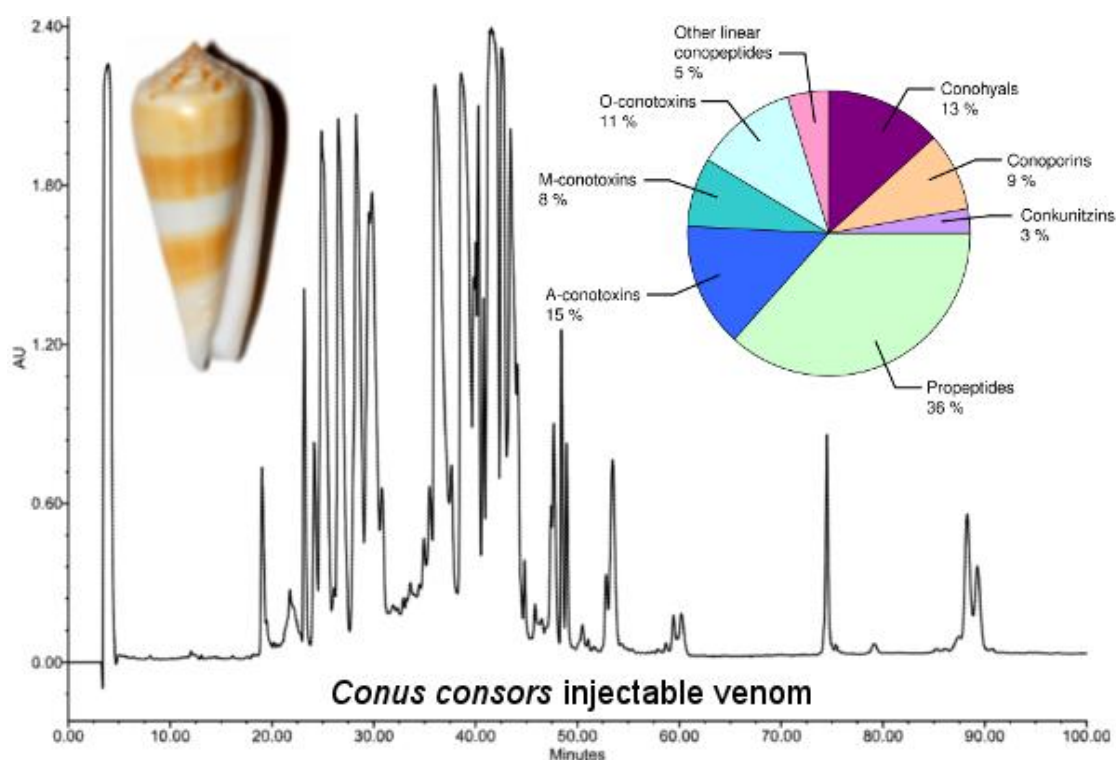
Conus venom, peptidomics, transcriptomics, conopeptides, conotoxins

Personal contribution

Part of the data analysis and manuscript proofreading.

Abstract

Predatory marine snails of the genus *Conus* use venom containing a complex mixture of bioactive peptides to subdue their prey. Here we report on a comprehensive analysis of the protein content of injectable venom from *Conus consors*, an indo-pacific fish-hunting cone snail. By matching MS/MS data against an extensive set of venom gland transcriptomic mRNA sequences, we identified 105 components out of ~400 molecular masses detected in the venom. Among them, we described new conotoxins belonging to the A, M- and O1-superfamilies as well as a novel superfamily of disulphide free conopeptides. A high proportion of the deduced sequences (36%) corresponded to propeptide regions of the A- and M-superfamilies, raising the question of their putative role in injectable venom. Enzymatic digestion of higher molecular mass components allowed the identification of new konkunitzins (~7 kDa) and two proteins in the 25 and 50 kDa molecular mass ranges respectively characterised as actinoporin-like and hyaluronidase-like protein. These results provide the most exhaustive and accurate proteomic overview of an injectable cone snail venom to date, and delineate the major protein families present in the delivered venom. This study demonstrates the feasibility of this analytical approach and paves the way for transcriptomics-assisted strategies in drug discovery.



1. Introduction

Cone snails from the genus *Conus* are carnivorous marine gastropods that use a complex venom apparatus to capture and digest their prey or to fight off predators and competitors. Their venoms have been extensively studied and are of particular interest because of the number and diversity of neuropharmacologically active peptides they contain. These peptides have been shown to target a wide range of therapeutically relevant receptors and ion-channels (Favreau & St"cklin, 2009). The peptide components of cone snail venom have attracted a lot of interest from the field of drug discovery (Gayler et al., 2005; Halai & Craik, 2009; Leary, Vierros, Hamon, Arico, & Monagle, 2009; Molinski, Dalisay, Lievens, & Saludes, 2009) and drug development (Favreau et al., 2012; Han, Teichert, Olivera, & Bulaj, 2008; Lewis, 2012). To date, one drug, derived from the venom of *C. magus*, has been approved to treat severe chronic pain (McGivern, 2007; Miljanich, 2004). With more than 700 *Conus* species already described and estimates ranging from 50 to more than 1000 venom components per species (Biass et al., 2009; Davis, Jones, & Lewis, 2009; Safavi-Hemami et al., 2011), it is difficult to predict the exact number of conotoxins that exist. Furthermore, several factors have been shown to influence the number and nature of molecules that can be identified in venoms. First, the method of ionisation (MALDI, ESI) and sample preparation (i.e. addition of a LC separation step) can significantly alter the resolution and detection of individual molecules, highlighting the need to use complementary techniques (Biass et al., 2009; Tayo, Lu, Cruz, & Yates, III, 2010). Second, the nature of the venom used as starting material for analysis (i.e. crude dissected venom or injectable venom, obtained by milking live animals and also named "injected venom" or "milked venom" in literature) can also significantly influence the overall results. Additionally, while interspecies variations have been well documented, recent findings have revealed large variations in venom composition between specimens of the same species (intraspecies variation), and even dramatic variations within a single individual over time (intraspecimen variation) (Davis et al., 2009;

Dutertre, Biass, Stocklin, & Favreau, 2010; Jakubowski, Kelley, Sweedler, Gilly, & Schulz, 2005).

Despite the variety in the characteristics of the mature peptides, conotoxins within a superfamily share typical cDNA coding sequences. These sequences usually encode one precursor (70–120 amino acids long) that is characterised by a gradient of divergence from the N- to the C-terminus. They exhibit a highly conserved signal sequence, followed by a less conserved propeptide region and finally a divergent hyper-variable mature toxin (except for a few conserved key residues such as cysteines involved in the formation of disulphide bonds) (Olivera et al., 1999). The evolution of the conotoxin gene family appears to be the result of gene duplications and high rates of non-synonymous substitutions, yet the detailed mechanisms underlying these genetic events are not fully understood (Conticello et al., 2001; Duda, Jr. & Palumbi, 1999).

This well-conserved and characteristic structure of conopeptide signal sequence is used to define the gene superfamilies. This allows targeted amplification of expressed cDNA sequence tags from venom ducts of different *Conus* species, revealing the expression profile of particular conotoxin superfamilies in the whole venom gland (Pi et al., 2006b; Pi et al., 2006a), as well as along the venom duct (Garrett, Buczek, Watkins, Olivera, & Bulaj, 2005). Recently, next generation sequencing technology has been used to unravel the complete venom gland transcriptomes of two fish hunting cone snails, *Conus bullatus* (Hu, Bandyopadhyay, Olivera, & Yandell, 2011) and *Conus consors* (Terrat et al., 2012). This comprehensive approach gives an overall view of the transcriptional activity in the venom duct, but does not clearly identify the components that are important for envenomation, (i.e., that actively contribute to the capture of the prey). Up to now, information on the global protein content of injected venom has been lacking, maybe because of the tedious milking procedure required to obtain the venom and the lack of genomic/transcriptomic data. Here, we report the composition of injectable venom from captive *C. consors*. We couple mass

spectrometry analysis to extensive venom duct transcriptomics and use bioinformatic tools to match the data and deduce nucleic and amino acid sequences. To our knowledge, this is the first time that such an approach has been used to characterise the conopeptide and conoprotein mixture injected into the prey.

2. Materials and methods

2.1 Chemicals

Acetonitrile (ACN, from Fisher Scientific Ltd., Loughborough, UK), trifluoroacetic acid (TFA, Thermo scientific, Rockford, IL, USA) and formic acid (FA, Acros Organics, Geel, Belgium) were of HPLC gradient grade or higher. Dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), iodoacetamide and all other reagents were of analytical grade or better and freshly prepared prior to use. Trypsin and chymotrypsin were sequencing grade (Worthington, Lakewood, NJ, USA). Purified deionised water was obtained using a Milli-Q system (Millipore Corp., Billerica, MA, USA). When required, solvents were filtered and sonicated before use.

2.2 Sample preparation

All specimens of *C. consors* used in this study were collected from one colony in the Chesterfield Islands (New Caledonia) in the frame of the CONFIELD-II scientific expedition in November 2008. Adults of similar size with no obvious external differences were kept in an aquarium and fed with fish. Injectable venom consisted in a pool of venom obtained regularly from some 35 specimens that were milked using a procedure adapted from Hopkins and colleagues (Hopkins et al., 1995) once a week over a few months period. Pools representing an average of 70 milked venoms were fractionated by RP-HPLC using an Alliance HT 2795 (Waters, Milford, MA, USA) separation module fitted with a Waters 996 Photodiode Array

Detector and operated with the Waters Millenium 4.0 software. Fractionation was performed using a 218TP54 C18 RP column (4.6×250 mm, 5 µm, Vydac, Hesperia, CA, USA). The flow rate was set to 0.8 mL/min with a gradient of 1% solvent B (90% ACN/0.1% TFA in water) per minute, starting from 100% of solvent A (0.1% TFA in water). Collected fractions were lyophilised and stored at –20 °C until use.

2.3 Venom duct transcriptome

First, 5 µg of total RNA was used to construct a cDNA library. RNA quality was assessed in a Bioanalyzer 2100 (Agilent-Bonsai Technologies). 5 µg of full-length double-stranded cDNA was then processed by the standard Genome Sequencer library-preparation method using the 454 DNA Library Preparation Kit (Titanium chemistry) to generate single-stranded DNA ready for emulsion PCR (emPCR™). The cDNA library was then nebulized according to the fragmentation process used in the standard Genome Sequencer shotgun library preparation procedure. The cDNA library was sequenced according to GS FLX technology (454/Roche). The short-reads from the sequencing were assembled by MIRA version 2.9.25 using enhanced 454 parameters.

2.4 Reduction of disulphide bonds

Lyophilised venom fractions were suspended in deionised water at an approximate concentration of 1 mg/mL (value based on the UV-HPLC chromatogram peak integration) and TCEP (50 mM in water, pH=4.5) was added to reach a concentration of 5 mM. The mixture was heated to 56 °C for 15 min then cooled and acidified to pH=3 with FA 20% prior to LC–ESI–MS and MS/MS analysis.

2.5 Proteolytic digestion experiments

Lyophilised venom fractions of interest were suspended in 30% ACN at an approximate concentration of 1 mg/mL (value based on the UV-HPLC peak integration). Disulphide bridges were reduced by incubating at 55 °C for 2 h in the presence of ammonium bicarbonate (2% w/v) and DTT 4.5 mM. Free sulfhydryl groups were then alkylated by addition of 10 µL iodoacetamide 100 mM (15 min, room temperature). Excess iodoacetamide was neutralised with 10 µL of cysteine 200 mM. Two separate digestions were carried out using trypsin and chymotrypsin, respectively. In both cases, 50 µL of bicarbonate 2% was added and enzyme concentration was adjusted to have an enzyme/protein ratio of 1:20 (w/w). Mixtures were then incubated at 37 °C for 6 h before acidification with 5 µL of FA (20%).

2.6. HPLC–ESI–MS and MS/MS analysis of venom fractions

Native injectable venom fractions were analysed by LC–ESI–MS. Reduced injectable venom fractions and digestion fragments were submitted to LC–ESI–MS with data directed acquisition of MS/MS. RP-HPLC was performed using an Alliance HT 2795 (Waters, Milford, MA, USA) separation module fitted with a Waters 996 Photodiode Array Detector and operated with the Waters Millennium 4.0 software. A Waters Atlantis T3 column (1×150 mm, 3 µm) was used for separation. The flow rate was 0.05 mL/min with a gradient of 1% solvent B (90% ACN/0.1% FA in water) per minute, starting from 100% of solvent A (0.1% FA in water). Blank runs confirmed there was no carry-over of peptides between runs. MS and MS/MS data acquisition was performed on a Q-TOF Ultima mass spectrometer (Waters-Micromass, Manchester, UK) equipped with its standard ESI source and operated in positive ionisation mode under control of the MassLynx 4.0 software with its Biolyux and MaxEnt3 options (Waters-Micromass). Molecular species were detected within the mass-to-charge ratio (m/z) range 100 to 1800. External calibration of the mass scale was performed with Glu-

fibrinopeptide. MS/MS survey was set up to adjust collision energy depending on the charge state of the fragmented ion, using automated charge state recognition. Four charge state ions were included in the survey.

2.7 MS and MS/MS data analysis

All MS data were manually interpreted using MassLynx and MS/MS data were automatically deconvoluted into singly-charged m/z species using Maxent3. Files were then submitted to the Phenyx software version 2.6.2 (Genebio, Geneva, Switzerland) and matched against the sequences deduced from the transcriptome database (all possible reading frames). Parent mass tolerance was set to 0.2 Da and the maximum p-value to 1.10^{-7} . Hydroxyproline, pyroglutamic acid, methionine oxidation and C-terminus amidation were included as possible modifications. For proteolytic fragments, main parameters were set with a maximum of two missed cleavages, and carboxyamidomethyl cysteine was included as fixed modification. Matches were considered valid when z-score was above 6 and delta m/z below 0.1. Matching results from MS and MS/MS with the corresponding transcript were then manually assessed and checked using Biolynx.

3. Results and discussion

3.1. Overview: mass range distribution and disulphide bridges

The venom batch was first fractionated by RP-HPLC and fractions exhibiting UV absorbance at 214 nm were selected for subsequent analysis (see chromatogram, Figure 29). Each native venom fraction was submitted to LC–ESI–MS analysis, which detected 419 unique molecular masses ranging from ~800 Da to ~50,000 Da. Almost 90% of these masses were between 1000 and 5500 Da and only 5% corresponded to large proteins (Figure 30A). Two large proteins had masses around 25 kDa and 50 kDa, respectively. Each fraction was then

chemically reduced and re-analysed by LC–ESI–MS following the same procedure. This enabled the identification 209 molecular masses that could be associated with their oxidised counter-parts, although some ambiguities remained due to possible partial reduction or incidental proximity of different masses. Most native molecular masses with no reduced counterpart were present in small amounts that did not allow further experimental work. Some non-associated reduced compounds may correspond to native dimers (Quinton, Gilles, & De, 2009), but no clear evidence of the presence of dimers could be found in this study. Traditionally, conopeptides are classified as “cysteine-rich” (also named conotoxins), containing 2 to 6 disulphide bridges or “cysteine-poor” containing one or no disulphide bridge (Norton & Olivera, 2006). In this study, 54% of the 209 peptides identified were cysteine-rich, most of them containing 2 or 3 disulphide bridges, and 46% were cysteine-poor (Figure 30B).

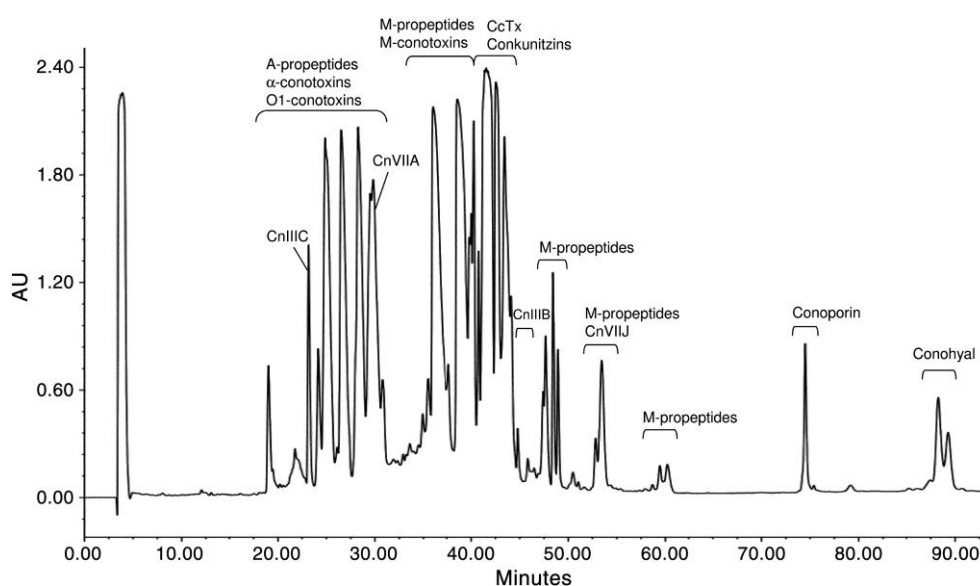


Figure 29: Typical UV (214 nm) chromatogram of pooled milked venom of *C. consors*.

The number of the latter is higher than expected when compared with entries in the ConoServer database (14%) (Kaas, Westermann, Halai, Wang, & Craik, 2008). It could be that past research efforts on conopeptides may have biased the search towards cysteine-rich

conopeptides (Ueberheide, Fenyo, Alewood, & Chait, 2009). Our results reveal that the relatively high number of cysteine-poor conopeptides is mainly consisting of propeptides (see next section), which has never been shown at the proteomic level prior to this study.

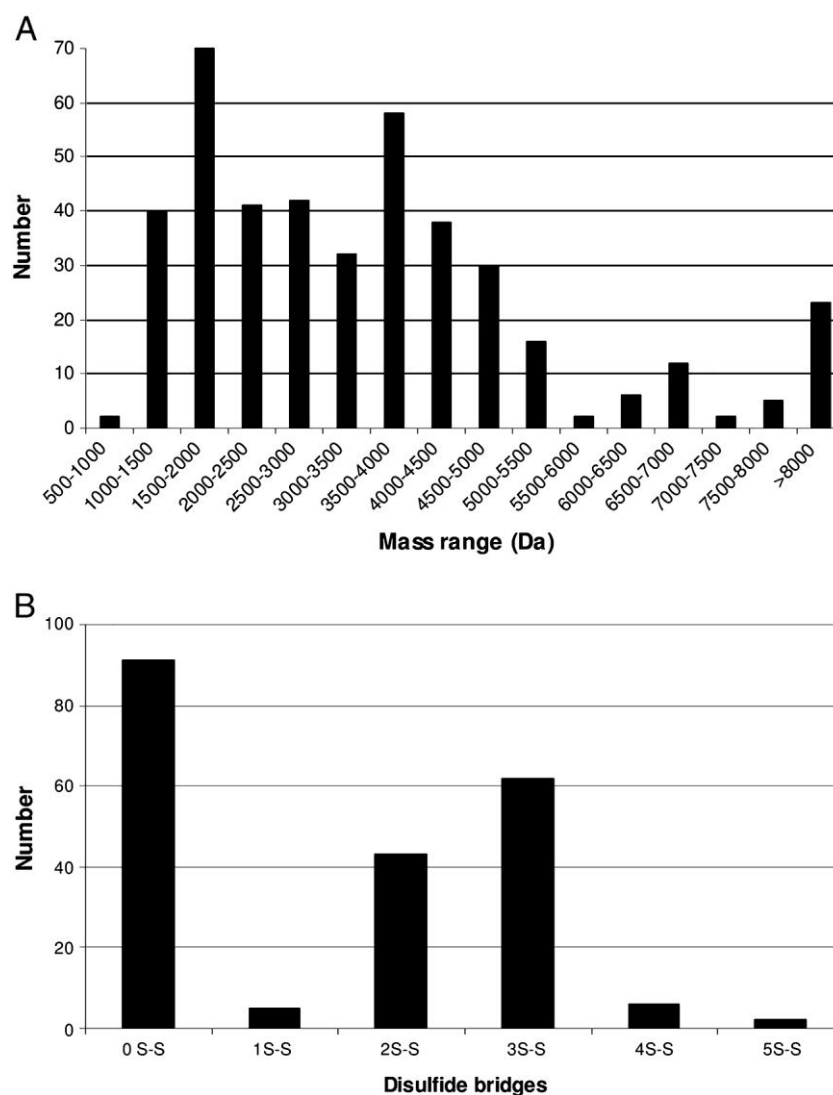


Figure 30: (A) Molecular mass range distribution of the 419 native compounds detected by ESI-MS analysis. (B) Number of disulphide bridges assigned for the 209 components for which a native and reduced mass could be determined without ambiguity.

3.2 Peptide identification

All reduced fractions were analysed by MS and MS/MS and data were matched against all possible amino acid sequences that could be deduced from an extensive database of the C.

consors venom gland transcriptome. A total of 79 sequences were identified using this automated approach (Table 7 to Table 12, see also Supplementary material 1). Among them, 16 conotoxins could be assigned to the A-superfamily, 8 to the M-superfamily, and 12 to the O1-superfamily. In a few cases, charge state recognition was not achieved properly and the data were re-analysed manually.

Table 7: A-conotoxins.

^a stands for C-ter amidation, O for hydroxyproline and S for glycosylated serine.

Name	Sequence	Observed native mass	Observed reduced mass	S–S
CnIA	GRCCHPACGKYYS ^a	1541.58	1545.72	2
CnIB	CCHPACGKYYS ^a	1328.47	1332.56	2
CnIC	CCHPACGKHF ^a	1286.48	1290.36	2
CnID	CCHPACGKHFN ^a	1313.48	1317.5	2
[Hyp4]-Cn ID	CCHOACGKHFN ^a	1329.5	1333.51	2
CnIE	CCHPACGGKYFK ^a	1410.56	1414.42	2
CnIF	GRCCHPACGKHF ^a	1499.58	1503.61	2
CnIG	CCHPACGGKYFK ^a	1353.54	1357.52	2
CnIH	NGRCCHPACGKHF ^a	1613.64	1617.65	2
[D1]-CnIH	DGRCCCHPACGKHF ^a	1614.63	1618.64	2
CnIJ	GRCCHPACGGKYFK ^a	1623.66	1627.57	2
CnIK	NGRCCHPACGKYYS ^a	1655.65	1659.87	2
[Hyp7]-CnIK	NGRCCHOACGKYYS ^a	1671.63	1675.81	2
CnIL	DGRCCCHPACGKYYS ^a	1656.6	1660.75	3
CcTx	AOWLVP <u>S</u> QITCCGYNOGTMCO ^a SCMCTNTC	4115.58	4121.7	3

Non-glycosylated CcTx	AOWLVP SQITTCGYN OGTCOSCMCTNTC	3223.09	3229.28	3
MGMRMMFTVFLLVVLTTTVVS FPSDSASDGRDDEAKDERSDIYESK RNGRCCHPACGKYYS CGRX α 3/5 CCHPACGKYYS C GRCCHPACGKYYS C NGRCCHPACGKYYS C NGRCCHOACGKYYS C 98_12623 CnIB CnIA CnIK [Hyp7]–CnIK				
MGMRMMFTVFLLVVLTTTVVS FPSDSASDV RDDEAKDERSD MYSK RNGRCCHPACGKHFSCGRX α 3/5 CCHPACGKHFSC GRCCHPACGKHFSC NGRCCHPACGKHFSC DGRCHPACGKHFSC 98_19786 CnIC CnIF CnIH [Asp1]–CnIH				
MGMRMMFTVFLLVVLTTTVVS FPSDSASDGRDDEAKDERSD MYELK RNGRCCHPACGKHFNCGRX α 3/5 CCHPACGKHFNC CCHOACGKHFNC 98_20026 CnID [Hyp4]–CnID				
MGMRMMFTVFLLVVLTTTVVS FPSDSASDGRDDEAKDERSDIYESK RDGRCHPACGKYYS CGRX α 3/5 DGRCHPACGKYYS C 100_30116 CnIL				
MMFTVFLLVVLTTTVVS FPSDSASDGRDDEAKDERSD MYELK RNGRCCHPACGGKYFKCGRX α 3/6 CCHPACGGKYFKC GRCCHPACGGKYFKC 98_6289 CnIE CnIJ				
LTTTVVS FPSDSASDGRDNEAKDERSD MYELK RNGRCCHPACGGKYFKCGRX α 3/5 CCHPACGGKYFKC 98_15425 CnIG				

Figure 31: α-Conotoxin contigs from which 14 mature sequences were identified in the injectable venom.

Loop lengths of the mature alpha-conotoxins are indicated on the left side of the mature parts. Signal and propeptide regions are shown in red and green, respectively.

In the A-superfamily, 14 sequences had the cysteine framework I and were similar to α-conotoxins that target muscular nicotinic acetylcholine receptors (Table 7). Two of these, CnIA and CnIB have been described previously (Favreau et al., 1999). All sequences derive from only 6 different contigs (Figure 31), exhibiting highly conserved signal and propeptide regions but allowing mature toxin diversification. This observation is in agreement with the findings of Santos et al. in 2004 (Santos, McIntosh, Hillyard, Cruz, & Olivera, 2004). From 6 precursors, mature toxin diversification through N-terminal drift in cleavage site and the introduction of PTMs (mainly hydroxyproline in the A-superfamily), resulted in 14 different conopeptides in the venom. Variants containing either asparagine or aspartic acid at the

same position were also observed (CnIH and CnIL), which could result from natural deamidation of the asparagine residue, a reaction that is well known in peptides, especially

Table 8 : M-conotoxins.

^a stands for C-ter amidation and Z for pyroglutamic acid.

Name	Sequence	Observed native mass	Observed reduced mass	S-S
CnIIIB	ZGCCGEPNLCFTRWCRNNARCCRQQ	2922.17	2928.2	3
[Gln1]-CnIIIB	QGCCGEPNLCFTRWCRNNARCCRQQ	2939.12	2945.22	3
CnIIIC	ZGCCNGPKGCSSKWCRDHARCC ^a	2373.85	2379.91	3
[Gln1]-CnIIIC	QGCCNGPKGCSSKWCRDHARCC ^a	2391	2397.01	3
CnIIID	RCC--RWPC-PR-KIDGEYCGCCL ^a	2352.86	2358.89	3
CnIIIE	RCCGEGASC-PVYSRDLICSCC	2373.97	2380	3
CnIIIF	RCCGEGASC-PRYFRNSQICSCC	2435.9	2441.92	3
CnIIIG	ZKCCGKGMTC-PRYFRDNFICGCC	2609.04	2615.2	3

when a glycine residue follows the asparagine (Tyler-Cross & Schirch, 1991) [32]. However, since one transcript containing the CnIL sequence was also identified, it remains to be determined whether CnIL is coded for, or results from deamidation of CnIK or a combination of both. It is also interesting to note that the cysteine pattern $\alpha 3/6$ is observed for the first time in CnIE and CnIJ, whereas all the other α -conotoxins structurally elucidated in this study shared a common $\alpha 3/5$ pattern.

Also belonging to the A-superfamily, the previously de-scribed CcTx venom component (see Ref. (Le et al., 1999) was identified in its non-glycosylated form as well as bearing the glycan group. The glycosylated sequence was first identified using an in-house bioinformatic tool already reported, and this work with full structural elucidation will be described elsewhere (Hocking et al. manuscript in preparation). Similar isoform diversity was observed among the

numerous conotoxins belonging to the M- and O1-superfamilies that were identified (Table 8 and Table 9), some of which had novel cysteine patterns. Of particular interest and observed for the first time in fish-hunting cone snails is a rare characteristic cysteine pattern shared by

Table 9 : O1-conotoxins.

^a stands for C-ter amidation, M for oxidised methionine, and O for hydroxyproline.

Name	Sequence	Observed native mass	Observed reduced mass	S–S
CnVIIA	CKGKGAOCTRLMYDCCHGSCSSSKGRC ^a	2846.14	2852.15	3
[Pro7]-CnVIIA	CKGKGAPCTRLMYDCCHGSCSSSKGRC ^a	2830.19	2836.22	3
[Pro7, oxMet12]-CnVIIA	CKGKGAPCTRLMYDCCHGSCSSSKGRC ^a	2846.14	2852.15	3
CnVIIB	CKGKGASCRRTSYDCCTGSCRS--GKC ^a	2609.02	2615.01	3
CnVIIC	CKGTGKPCSRIAYNCCTGSCRS--GKC ^a	2575.05	2581.41	3
[Hyp7]-CnVIIC	CKGTGKOCsRIAYNCCTGSCRS--GKC ^a	2591.03	2597.22	3
CnVIID	CKGKGASCSRTMYNCCSGSC--NRGKC ^a	2595.99	2602.03	3
CnVIIE	CKGTGKOCsRIAYNCCTGSCRS--GKCG	2649.06	2655.23	3
CnVIIF	CKGKGASCSRTMYNCCSGSC--NRGKCG	2653.96	2660.05	3
CnVIIG	CKGKGASCRRTSYDCCTGSCRS--GKCG	2667.03	2673.12	3
CnVIIH	CKGKGAOCTRLMYDCCHGSCSSSKGRCG	2904.17	2910.17	3
CnVIIJ	STSCMKAGSYCRSTTRTCC-GYCAYFGKFCIDFPSN	3851.49	3857.59	3

some sequences of the M-superfamily, such as CnIIID-G, where a single residue stands in loop 4, before the last cysteine pair (Han et al., 2006; Wang et al., 2009). Despite showing a paralytic effect in fish (SD and PF, personal observation), the molecular target of these peptides is currently unknown. In the O1-superfamily, most isoforms shared an identical cysteine framework with a final loop that varied in length (4 or 6 amino acids). In this isoform

group, positions 13 and 2 always consists in a Tyr and a Lys, respectively, both of these have been shown to be essential for the interaction with voltage-sensitive Ca^{2+} channels (Lew et al., 1997). Of note, CnVIIJ is a unique sequence that does not belong to the previous group due to a divergent amino acid sequence and atypical sequence length.

Table 10: A-propeptides.

O stands for hydroxyproline.

Name	Sequence	Observed native mass	Observed reduced mass	S–S
A-propeptide-Cn1	FPSDSASDGRD	1152.52	1152.5	0
A-propeptide-Cn2	FPSDSASDGRDD	1267.54	1267.41	0
[Hyp2]-A-propeptide-Cn2	FOSDSASDGRDD	1283.48	1283.6	0
A-propeptide-Cn3	FPSDSASDVRRD	1309.52	1309.7	0
A-propeptide-Cn4	FPSDSASDGRDDE	1396.54	1396.54	0
A-propeptide-Cn5	FPSDSASDGRDDEA	1467.61	1467.7	0
[Hyp2]-A-propeptide-Cn5	FOSDSASDGRDDEA	1483.57	1483.59	0
A-propeptide-Cn6	FPSDSASDGRDDEAK	1595.65	1595.58	0
[Hyp2]-A-propeptide-Cn6	FOSDSASDGRDDEAK	1611.64	1611.65	0
A-propeptide-Cn7	FPSDSASDGRDDEAKDE	1839.74	1839.78	0
[Hyp2]-A-propeptide-Cn7	FOSDSASDGRDDEAKDE	1855.64	1855.69	0
A-propeptide-Cn8	FPSDSASDVRRDDEAKDE	1881.8	1881.75	0
A-propeptide-Cn9	FPSDSASDGRDDEAKDER	1995.86	1995.88	0
A-propeptide-Cn10	OLVVSFPSDSASDGRDDEAKDE	2351.01	2350.97	0

Table 11: M-propeptides.

M stands for oxidised methionine and O for hydroxyproline.

Name	Sequence	Observed native mass	Observed reduced mass	S-S
M-propeptide-Cn1	SEQYPLFDM	1128.5	1128.49	0
M-propeptide-Cn2	LPMDGDQPADRPA	1381.67	1381.63	0
M-propeptide-Cn3	DISSEQYPLFDM	1443.62	1443.62	0
[oxMet11]-M-propeptide-Cn3	DISSEQYPLFDM	1459.62	1459.62	0
M-propeptide-Cn4	ZDDISSEKHPLFD	1512.6	1512.67	0
M-propeptide-Cn5	DDISSEQYPLFDM	1558.65	1558.68	0
[oxMet13]-M-propeptide-Cn5	DDISSEQYPLFDM	1574.67	1574.68	0
M-propeptide-Cn6	ZDDISSEQHPLFNQ	1639.72	1639.61	0
M-propeptide-Cn7	LDDISSEQYPLFDM	1671.72	1671.77	0
[oxMet14]-M-propeptide-Cn7	LDDISSEQYPLFDM	1687.73	1687.78	0
M-propeptide-Cn8	MLDDISSEQYPLFDM	1802.76	1802.77	0
[oxMet15]-M-propeptide-Cn8	MLDDISSEQYPLFDM	1818.77	1818.77	0
[oxMet1,15]-M-propeptide-Cn8	MLDDISSEQYPLFDM	1834.77	1834.75	0
M-propeptide-Cn9	MLDDISSEQYPLFDMRKSAATA	2586.17	2586.31	0
M-propeptide-Cn10	LPMDDEDQPADQPADRMQDDISSEQYPLFD	3365.45	3365.48	0
[oxMet16]-M-propeptide-Cn10	LPMDDEDQPADQPADRMQDDISSEQYPLFD	3381.32	3381.46	0
[Hyp2,12]-M-propeptide-Cn10	LOMDDEDQPADQPADRMQDDISSEQYPLFD	3397.31	3397.48	0
[oxMet3,Hyp12]-M-propeptide-Cn10	LPMDDEDQPADQPADRMQDDISSEQYPLFD	3397.31	3397.48	0
M-propeptide-Cn11	LPMDDEDQOAEFADRMMLDDISSEQYPLFDM	3480.5	3480.65	0
[Pro8,Hyp26]-M-propeptide-Cn11	LPMDDEDQPAEFADRMMLDDISSEQYOLFD	3480.53	3480.66	0
[oxMet30]-M-propeptide-Cn11	LPMDDEDQOAEFADRMMLDDISSEQYPLFDM	3496.5	3496.58	0
[Pro8,oxMet3,16]-M-propeptide-Cn11	LPMDDEDQPAEFADRMMLDDISSEQYPLFDM	3496.5	3496.58	0
M-propeptide-Cn12	FPMDEDQPAEFADRMMLDDISSEQYPLFDM	3498.54	3498.64	0
[Hyp2]-M-propeptide-Cn12	FOMDEDQPAEFADRMMLDDISSEQYPLFDM	3530.52	3530.61	0

Unexpectedly, 38 propeptides belonging to the A- and M-superfamilies were also identified (Table 10 and Table 11). Unlike the O-superfamily propeptides, these were Arg- and Lys-poor, which is likely to make them less susceptible to degradation by endoproteases. In the case of the O-superfamily propeptides, the resulting enzymatic digest would give rise to peptide fragments with low molecular masses (<400 Da), which are unlikely to be detected by our analytical approach. It is not clear whether these A- and M-superfamily propeptides are by-products of incomplete enzymatic degradation of the precursor, or if they have a biological function on their own. Although most propeptides could be associated with mature toxins detected in the injectable venom, two linear peptides were unambiguously identified as M-superfamily propeptides; however, these could not be matched to any contigs coding for an associated mature toxin (Figure 32). This is rather intriguing, and while these contigs may be remnants or evolution of a full precursor with a mature toxin, the identification of PTMs (oxidised methionine, pyroglutamic acid and hydroxyproline residues) on a large proportion of all propeptides detected (53%) could support the hypothesis of a specific role for these propeptides. Yet, this hypothesis still has to be investigated.

```

98_ 6246  MSKLGVLLTICLLLFPLTALPMDGDSVDRPAERMQDDISSEQYPLFNQKRRCCGEGASCPVYSRDRLICSCCX
98_ 6511  MSKLGVLLTICLLLFPLTALPMDDEDQPADQPADRMQDDISSEQYPLFDKRQKCCGKGMTCPRYFRDNFICGCCX
98_ 6542  MSKLGVLLTICLLLFPLTALPMDDEDQPAPEPADRMLDDISSEQYPLFDMRKSAAAVSETVX
98_ 6543  MSKLGVLLTICLLLFPLTALPMDDEDQPAPEPADRMLDDISSEQYPLFDMRKSAAAKAX
98_18606 MSKLGVLLTICLLLFPLTALPMDDEDQPAPEPADRMLDDISSEQYPLFDMRKSTAAKAX

```

Figure 32: Example of M-superfamily contig alignment with propeptide and mature toxins identified in the injectable venom.

Contigs 98_6246 and 98_6511 are precursors of CnIII E and CnIII G, respectively. Signal and propeptide regions are shown in red and green, respectively. Peptides identified in the venom by LC–ESI–MS/MS are underlined.

Five peptides containing no cysteines were found in the same transcript and could not be associated with any known superfamily (Table 12 and Figure 33). A BLAST search of this

transcript against the Uniprot database resulted in one hit (Q2HZ30) with an e-value of 1.84×10^{-37} . The identified high frequency protein 1 is found in the *Conus litteratus* cDNA library but, to date, has not been identified at the protein level (Pi et al., 2006a). Alignment of both sequences showed a well-conserved signal region that did not share any significant similarity with other known superfamilies. Although the activity or function of these expressed peptides is still unknown, they may represent a new superfamily of disulphide-poor conopeptides.

Table 12 : Other linear conopeptides.

M stands for oxidised methionine.

Name	Sequence	Observed native mass	Observed reduced mass	S-S
Linear conopeptide-Cn1	DPALMSMQGMQGGQMPGMAGGQ	2178.89	2178.92	0
Linear conopeptide-Cn2	FDPALMSMQGMQGGQMPGMAGGQ	2325.88	2325.97	0
[oxMet6]-Linear conopeptide-Cn2	FDPAL <u>MS</u> MQGMQGGQMPGMAGGQ	2341.9	2342	0
Linear conopeptide-Cn3	NADENKAPFDSEENFMN	1970.76	1970.79	0
Linear conopeptide-Cn4	KQHSQFNADENKAPFDSEENFMN	2726.09	2726.15	0

C. litteratus MLRLII-AAVLASACLAYPAKRDGAPADSANLQSFPGMMQAMPNMQGMQPMQAMPAGQFMPFNPNLGIGYKRAVYE
C. consors MLRLIIAAAVLVSACLAYPQRREGAPADAANLQSFDPALM-SMQGMQGGQMP--GMAGGQFLPFNPNLQMGYKRDFDE
 ***** :*:*****:*****.*:* :* .*** * * .*.***:***** :**** .**
C. litteratus DREERRPHSKFN--DNKSPFEQEGGLGNFMNFM--KENGNLFPANMDGAAADLGKFEPKSAEKGQKEGQFRFFDKQQ--
C. consors NLEKRKQHSQFNADENKAPFDSE---ENFMNFLHNEKGDKHPFANVDSADTDLGQFEPKSAE--NKNGEFRFFDKQEX
 :*:*: **:*** :*:***:* *****: :*:*: *****:* :*:*****:*

Figure 33: Alignment of linear sequence from *C. litteratus* (Q2HZ30) and *C. consors*.

Signal sequences are shown in red. Peptides identified in the venom by LC-ESI-MS/MS are underlined.

3.3. Elucidation of mini-protein and protein primary structures

Components above 4000 Da were not identified automatically due to poor collision-induced dissociation fragmentation of larger molecules. HPLC fractions containing high molecular mass compounds were therefore submitted to reduction, alkylation and selective proteolytic degradation by trypsin and chymotrypsin. The proteolytic fragments were then analysed by LC–MS/MS and data were matched against the venom duct transcriptome database. Three molecules in the 7 kDa molecular mass range were identified as conkunitzins, namely Conk-Cn1 (90% coverage), Conk-Cn2 (80% coverage) and Conk-Cn3 (41% coverage). These data will be presented in detail in an article dedicated to Conkunitzins of *C. consors* venom. Conkunitzins have already been described in venom from the genus *Conus*, in particular Conk-S1 has been identified in venom from *Conus striatus* (Bayrhuber et al., 2005). These proteins are known for their activity on potassium channels and the three isoforms identified here are currently under further investigation.

It has already been reported that injectable venom and dissected venom exhibit different HPLC profiles (Biass et al., 2009). In particular, injectable venom contains less hydrophobic compounds, which was confirmed in this study, with no δ -conotoxins detected and only very few proteins. Nevertheless, two proteins in the 25 and 50 kDa molecular mass ranges were identified in the injectable venom. The 50 kDa protein corresponded to hyaluronidase-like protein, with a major isoform named Conohyal-Cn1 (Violette et al., 2012). In total, 14 molecular masses were assigned to hyaluronidases, a diversity mainly created by only a few isoforms with varying glycosylation patterns. Hyaluronidases are well known to enhance diffusion of the venom by degrading the extracellular matrix (Stern & Jedrzejewski, 2006). In the 25 kDa molecular mass range, three digest fragments allowed the unambiguous identification of a full-length nucleotide sequence coding for an actinoporin-like protein, named Conoporin-Cn1. Figure 34 presents an alignment of Conoporin-Cn1 with echotoxin-2, isolated from the marine gastropod, *Monoplex echo* (Kawashima, Nagai, Ishida, Nagashima,

& Shiomi, 2003); both sequences showed 36% homology. A total of 9 isoforms were detected in the injectable venom, mainly corresponding to different oxidative states. Echotoxins are monomeric and cysteine-free proteins belonging to the sea anemone actinoporin family, which is known for its haemolytic properties (Shiomi, Kawashima, Mizukami, & Nagashima, 2002). These two protein families, hyaluronidases and actinoporins, have a clearly de-fined functional role and their presence in the injected venom appears to be fully justified. On the other hand, enzymes involved in the synthesis and maturation of conotoxins that were previously identified in dissected venom ducts and which were also found in the *C. consors* venom gland transcriptome were not detected in the injectable venom, as might be expected. This strongly supports the existence of a protein selection mechanism along or at the distal part of the venom duct, as previously suggested (Tayo et al., 2010).

```

M. echo      MKRNLAL----VVVVALISQSRPAESAGGTIIATLSKIPLSTLASALNTALETGASVASAAAA-ATSSDY
C. consors   MGVOFPALKTMVTVFLLLMGNMSP-----VVMKSSIPLRKVKMVASKVVPFGSSLYGVALKDLADTSY
              *  :: **      .*: *: : *      :.  *.***  .:  . . . . :*::: .*      :.:*

M. echo      SVTCVIEVENWTKHLMKYPVVQIANSGGLLTVAKNVLPAEIQSFAMRK-AWGANGVYGTVSWVLGQTNRNV
C. consors   NITCTLQVENWIRYRLIVPSVQMV-YGVVTTTPIAIEPAKREAFAVRKTSDTASGVAGSVSWELEKARRRF
              .:***.:***: :: : * **: . * : *.. : **: :***:*** : *.** *:*** * :.:*.

M. echo      VIMWSAPYNFDFYSNWLAVGMSRPGLAVPSSRSTWFDLMYYGNSNADISFCTRRILPTVSTQSI SRTLSGK
C. consors   VIMWSVPDNFNSFGYWMGLGMTREGLVDPD--KDWYGQMYSGSSDGLTF-TRKDFS YNTDSIIYSNDKFE
              ***** * **: . . *: :*: * ** . * . * :. ** *: :*: * **: . . : . * . . :

M. echo      SEGSMNNIHKARVRATVKPIKT--MDLASSILTKLEALAGANGKRAIQQELARRAEEEEKQRKRKALDEMLK
C. consors   VEGDMTNTQHAQIKIVIRPSSNNWKDLAPKIRRKLKKKPKPARQR-----DN--X
              **.*. * :*: :. :*: . . ***** * **: . . :*

```

Figure 34: Alignment of actinoporin-like sequences from *M. echo* (Q76CA2) and *C. consors* (Conoporin-Cn1).

Signal sequences are shown in red. Underlined fragments were those identified from a tryptic and chymotryptic digest.

3.4 Molecular overview of injectable venom

This large-scale identification of peptides and proteins in injectable venom from *C. consors* and elucidation of their sequence represents a significant advance, as there is presently little detailed information on these complex mixtures (Biass et al., 2009; Jakubowski et al., 2005). The combined approach of venom proteomics and venom gland transcriptomics described here offers several advantages for the study of venoms. First, a very small amount of venom sample is required, and there is no need for isolation and purification of the different components. Second, the automatic ESI–MS/MS acquisition is rapid and efficient. Even if fragmentation parameters are not perfectly optimised, the data allowed the identification of 25% of the detected masses, containing a high proportion of propeptides (36%) and conopeptides (39%) (Figure 35). The recovery and identification of unassigned molecular masses (about 300) in the injectable venom represents a major challenge, and additional analytical techniques will be required to identify peptides present in small amounts. This semi-automated approach is limited by two factors: the lack of fragmentation of sequences above 4000 Da, which could be partially overcome by using further enzymatic degradation; and the complexity of post-translational modifications in *Conus* venoms (Craig, Bandyopadhyay, & Olivera, 1999), which was partially solved by manual interpretation and use of an in-house bioinformatic tool. However, the high number and diversity of PTMs that can be combined in *Conus* venom peptides will most likely continue to present a challenge for future proteomic studies.

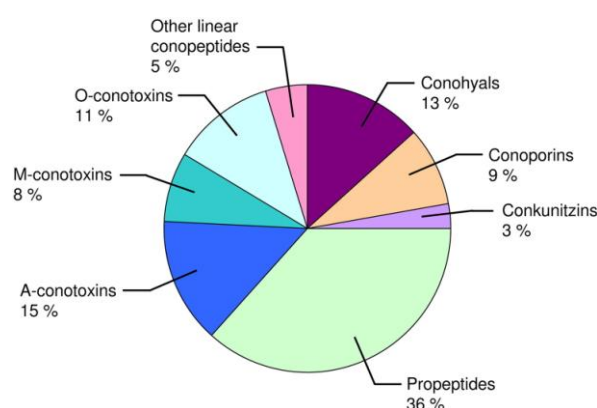


Figure 35: Family distribution of identified conopeptides and conoproteins in *C. consors* injectable venom.

3.5 Functional overview of identified compounds

This global proteomic picture of injectable venom allows us to infer the physiological mechanisms underlying the rapid capture of prey by *C. consors*. First, the immediate tetanic paralysis observed during prey envenomation is likely due to the action of the major peptide CcTx, which has been reported to activate Na⁺ channels on axons and motor nerve terminals (Le et al., 1999). This rapidly and efficiently incapacitates the prey, but its effect is short-lasting. The α -, μ - and ω -conotoxins prolong the paralysis by disrupting signal transduction by blocking nicotinic acetylcholine receptors, and voltage-gated Na⁺ and Ca²⁺ channels. The paralysed prey can then be engulfed safely, without risk of injuries from a struggling fish. The newly identified Conohyals and Conoporins most likely participate in the envenomation process by disrupting the cellular and extracellular matrices of the prey to favour rapid diffusion of other venom bioactives, thus playing a key role in the envenomation process. Our findings also raise questions about the function of propeptides and disulphide-free conopeptides in the venom. It will certainly be interesting to test the hypothesis of their potential biological significance by extensive pharmacological studies on these compounds. The results of this model study demonstrate the validity of a combined proteomic and

transcriptomic approach to identify novel peptides in *Conus* venoms, which can be exploited for the accelerated identification of bioactive ingredients in drug discovery.

3.6 Concluding remarks

The present study shows that it is possible to identify more than 100 peptides and proteins in injectable cone snail venom using a combined proteomic/transcriptomic approach, and establishes this as a promising strategy for the analysis of venoms. Nevertheless, a number of native molecular masses could not be identified using this technique, and follow-up investigations will involve a more in-depth characterisation of these peptides and proteins. Progress in mass spectrometry, especially in sensitivity, accuracy and speed, when coupled to massive NextGen genome and transcriptome sequencing, will improve component identification. The precise and global composition of a venom inferred from such analyses clearly provides better insights into the envenomation process for prey capture. Most importantly, this detailed venom proteomic panorama lays the basis to understand the numerous chemical and biological mechanisms in action in the venom apparatus to produce such efficient bioactive ingredients.

Acknowledgements

We would like to express our deepest gratitude to the Governments of New Caledonia and French Polynesia, the French Navy, the IRD-Nouméa “Institut de Recherche pour le Développement” (Fabrice Colin, Jean-Louis Menou, Claude Payri, Napoléon Colombani) as well as to the Toxinomics Foundation office in Nouméa (Alain Gerbault and Jacques Pusset) for their constant support. We are most grateful to the European Commission for financial strategy and scientific support: this study has been performed as part of the CONCO cone snail genome project for health (www.conco.eu) within the 6th Framework Program (LIFESCIHEALTH-6 Integrated Project LSHB-CT-2007, contract number 037592) with Dr

Torbjörn Ingemansson as scientific officer. Last but not least, we would like to express our gratitude to Robin Offord and Renée Ménez who participated to the sample collection missions CONFIELD-I in 2007 and CONFIELD-II in 2008, respectively. Finally we thank our colleagues at Atheris Laboratories Cécile Cros, Frédéric Perret and Nicolas Puillandre for their help.

References

- Bayrhuber, M., Vijayan, V., Ferber, M., Graf, R., Korukottu, J., Imperial, J. et al. (2005). Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. *J.Biol.Chem.*, 280, 23766-23770.
- Bias, D., Dutertre, S., Gerbault, A., Menou, J. L., Offord, R., Favreau, P. et al. (2009). Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J.Proteomics.*, 72, 210-218.
- Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., & Fainzilber, M. (2001). Mechanisms for evolving hypervariability: the case of conopeptides. *Mol.Biol.Evol.*, 18, 120-131.
- Craig, A. G., Bandyopadhyay, P., & Olivera, B. M. (1999). Post-translationally modified neuropeptides from *Conus* venoms. *Eur.J.Biochem.*, 264, 271-275.
- Davis, J., Jones, A., & Lewis, R. J. (2009). Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides*, 30, 1222-1227.
- Duda, T. F., Jr. & Palumbi, S. R. (1999). Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc.Natl.Acad.Sci.U.S.A*, 96, 6820-6823.
- Dutertre, S., Bias, D., Stocklin, R., & Favreau, P. (2010). Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon*, 55, 1453-1462.
- Favreau, P., Benoit, E., Hocking, H. G., Carlier, L., D', H. D., Leipold, E. et al. (2012). A novel micro-conopeptide, CnIIIC, exerts potent and preferential inhibition of Na(V) 1.2/1.4

channels and blocks neuronal nicotinic acetylcholine receptors. *Br.J.Pharmacol.*, 166, 1654-1668.

Favreau, P., Krimm, I., Le, G. F., Bobenrieth, M. J., Lamthanh, H., Bouet, F. et al. (1999). Biochemical characterization and nuclear magnetic resonance structure of novel alpha-conotoxins isolated from the venom of *Conus consors*. *Biochemistry*, 38, 6317-6326.

Favreau, P. & Stöcklin, R. (2009). Marine snail venoms: use and trends in receptor and channel neuropharmacology. *Curr.Opin.Pharmacol.*, 9, 594-601.

Garrett, J. E., Buczek, O., Watkins, M., Olivera, B. M., & Bulaj, G. (2005). Biochemical and gene expression analyses of conotoxins in *Conus textile* venom ducts. *Biochem.Biophys.Res.Comm.*, 328, 362-367.

Gayler, K., Sandall, D., Greening, D., Keays, D., Polidano, M., Livett, B. et al. (2005). Molecular prospecting for drugs from the sea. Isolating therapeutic peptides and proteins from cone snail venom. *IEEE Eng Med.Biol.Mag.*, 24, 79-84.

Halai, R. & Craik, D. J. (2009). Conotoxins: natural product drug leads. *Nat.Prod.Rep.*, 26, 526-536.

Han, T. S., Teichert, R. W., Olivera, B. M., & Bulaj, G. (2008). *Conus* venoms - a rich source of Peptide-based therapeutics. *Curr.Pharm.Des*, 14, 2462-2479.

Han, Y. H., Wang, Q., Jiang, H., Liu, L., Xiao, C., Yuan, D. D. et al. (2006). Characterization of novel M-superfamily conotoxins with new disulfide linkage. *FEBS J.*, 273, 4972-4982.

- Hopkins, C., Grilley, M., Miller, C., Shon, K. J., Cruz, L. J., Gray, W. R. et al. (1995). A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J.Biol.Chem.*, 270, 22361-22367.
- Hu, H., Bandyopadhyay, P. K., Olivera, B., & Yandell, M. (2011). Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC.Genomics*, 12, 60.
- Jakubowski, J. A., Kelley, W. P., Sweedler, J. V., Gilly, W. F., & Schulz, J. R. (2005). Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J.Exp.Biol.*, 208, 2873-2883.
- Kaas, Q., Westermann, J. C., Halai, R., Wang, C. K., & Craik, D. J. (2008). ConoServer, a database for conopeptide sequences and structures. *Bioinformatics.*, 24, 445-446.
- Kawashima, Y., Nagai, H., Ishida, M., Nagashima, Y., & Shiomi, K. (2003). Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. *Toxicon*, 42, 491-497.
- Le, G. F., Favreau, P., Benoit, E., Mattei, C., Bouet, F., Menou, J. L. et al. (1999). A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na⁺-dependent mechanism. *Eur.J.Neurosci.*, 11, 3134-3142.
- Leary, D., Vierros, M., Hamon, G. I., Arico, S., & Monagle, C. (2009). Marine genetic resources: A review of scientific and commercial interest. *Marine Policy*, 33, 183-194.
- Lew, M. J., Flinn, J. P., Pallaghy, P. K., Murphy, R., Whorlow, S. L., Wright, C. E. et al. (1997). Structure-function relationships of omega-conotoxin GVIA. Synthesis, structure, calcium channel binding, and functional assay of alanine-substituted analogues. *J.Biol.Chem.*, 272, 12014-12023.

- Lewis, R. J. (2012). Discovery and development of the chi-conopeptide class of analgesic peptides. *Toxicon*, 59, 524-528.
- McGivern, J. G. (2007). Ziconotide: a review of its pharmacology and use in the treatment of pain. *Neuropsychiatr.Dis.Treat.*, 3, 69-85.
- Miljanich, G. P. (2004). Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr.Med.Chem.*, 11, 3029-3040.
- Molinski, T. F., Dalisay, D. S., Lievens, S. L., & Saludes, J. P. (2009). Drug development from marine natural products. *Nat Rev.Drug Discov.*, 8, 69-85.
- Norton, R. S. & Olivera, B. M. (2006). Conotoxins down under. *Toxicon*, 48, 780-798.
- Olivera, B. M., Walker, C., Cartier, G. E., Hooper, D., Santos, A. D., Schoenfeld, R. et al. (1999). Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann.N.Y.Acad.Sci.*, 870, 223-237.
- Pi, C., Liu, J., Peng, C., Liu, Y., Jiang, X., Zhao, Y. et al. (2006a). Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics*, 88, 809-819.
- Pi, C., Liu, Y., Peng, C., Jiang, X., Liu, J., Xu, B. et al. (2006b). Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie*, 88, 131-140.
- Quinton, L., Gilles, N., & De, P. E. (2009). TxXIIIA, an atypical homodimeric conotoxin found in the *Conus textile* venom. *J.Proteomics.*, 72, 219-226.
- Safavi-Hemami, H., Siero, W. A., Gorasia, D. G., Young, N. D., Macmillan, D., Williamson, N. A. et al. (2011). Specialisation of the venom gland proteome in predatory cone snails

- reveals functional diversification of the conotoxin biosynthetic pathway. *J.Proteome.Res.*, 10, 3904-3919.
- Santos, A. D., McIntosh, J. M., Hillyard, D. R., Cruz, L. J., & Olivera, B. M. (2004). The A-superfamily of conotoxins: structural and functional divergence. *J.Biol.Chem.*, 279, 17596-17606.
- Shiomi, K., Kawashima, Y., Mizukami, M., & Nagashima, Y. (2002). Properties of proteinaceous toxins in the salivary gland of the marine gastropod (*Monoplex echo*). *Toxicon*, 40, 563-571.
- Stern, R. & Jedrzejewski, M. J. (2006). Hyaluronidases: their genomics, structures, and mechanisms of action. *Chem.Rev.*, 106, 818-839.
- Tayo, L. L., Lu, B., Cruz, L. J., & Yates, J. R., III (2010). Proteomic analysis provides insights on venom processing in *Conus textile*. *J.Proteome.Res.*, 9, 2292-2301.
- Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stocklin, R. et al. (2012). High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon*, 59, 34-46.
- Tyler-Cross, R. & Schirch, V. (1991). Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. *J.Biol.Chem.*, 266, 22549-22556.
- Ueberheide, B. M., Fenyo, D., Alewood, P. F., & Chait, B. T. (2009). Rapid sensitive analysis of cysteine rich peptide venom components. *Proc.Natl.Acad.Sci.U.S.A*, 106, 6910-6915.

- Violette, A., Leonardi, A., Piquemal, D., Terrat, Y., Biass, D., Dutertre, S. et al. (2012). Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar. Drugs*, 10, 258-280.
- Wang, L., Liu, J., Pi, C., Zeng, X., Zhou, M., Jiang, X. et al. (2009). Identification of a novel M-superfamily conotoxin with the ability to enhance tetrodotoxin sensitive sodium currents. *Arch. Toxicol.*, 83, 925-932.

6.5 Concluding remarks

This work represents the first global proteomic analysis of cone snail injected venom ever achieved through transcriptomic data match with MS/MS experimental spectra. Although the identification covered only about 25% of the detected masses, the transcriptomic and proteomic data combination strategy proved to be quite successful as it led to the rapid identification of more than 100 components, many bearing PTMs, in the injectable venom of a cone snail.

Not only did these sequence characterizations enrich the venom peptide databases for new potential drug candidates, they also provided a great functional view of the cocktail required by the envenomation process that occurs in prey capture. Indeed, it gave good insights into the physiological mechanisms used by the piscivorous *C. Consors* for swift prey capture. As an example, the major conopeptide – the CcTx, a voltage-gated sodium channel activator¹⁷ – present in the injectable venom provokes a very rapid excitotoxic response in the prey, and efficient paralysis. However, the effect of the CcTx is not long lasting, and it is followed by the effect of the α -, μ - and ω -conotoxins found in the cocktail. These take over the paralysis action by blocking nACh receptors, sodium and calcium voltage-gated channels. Mini-proteins belonging to the conkunitzins family, as well as two groups of proteins respectively belonging to hyaluronidases and porins, were also found in the injectable venom mixture. The role of these proteins is well defined as they are known to enhance the diffusion of the venom within the prey. Their presence is confirmed in the following study, which precisely focuses on high molecular mass proteins.

***Conus consors* snail venom proteomics proposes functions, pathways, and novel families involved in its venom system**

Adrijana Leonardi¹, Daniel Biass², Dusan Kordis¹, Reto Stocklin², Philippe Favreau², Igor Krizaj^{1,3,4}

¹ Department of Molecular and Biomedical Sciences, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

² Atheris Laboratories, Case postale 314, CH-1233 Bernex-Geneva, Switzerland

³ Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Askerceva 5, SI-1000 Ljubljana, Slovenia

⁴ Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova 39, SI-1000 Ljubljana, Slovenia

Publication data

Published the 29.08.2012 in the Journal of Proteome Research, Vol. 10, Issue 11, p. 5046-5058.

Pubmed ID: 22928724

DOI: 10.1021/pr3006155

Cited 13 times on Google Scholar.

Keywords

Cone snail, *Conus consors*, venom, two-dimensional gel electrophoresis, mass spectrometry, proteomics, venomics, post-translational modification, selection process, conoprotein, conotoxin

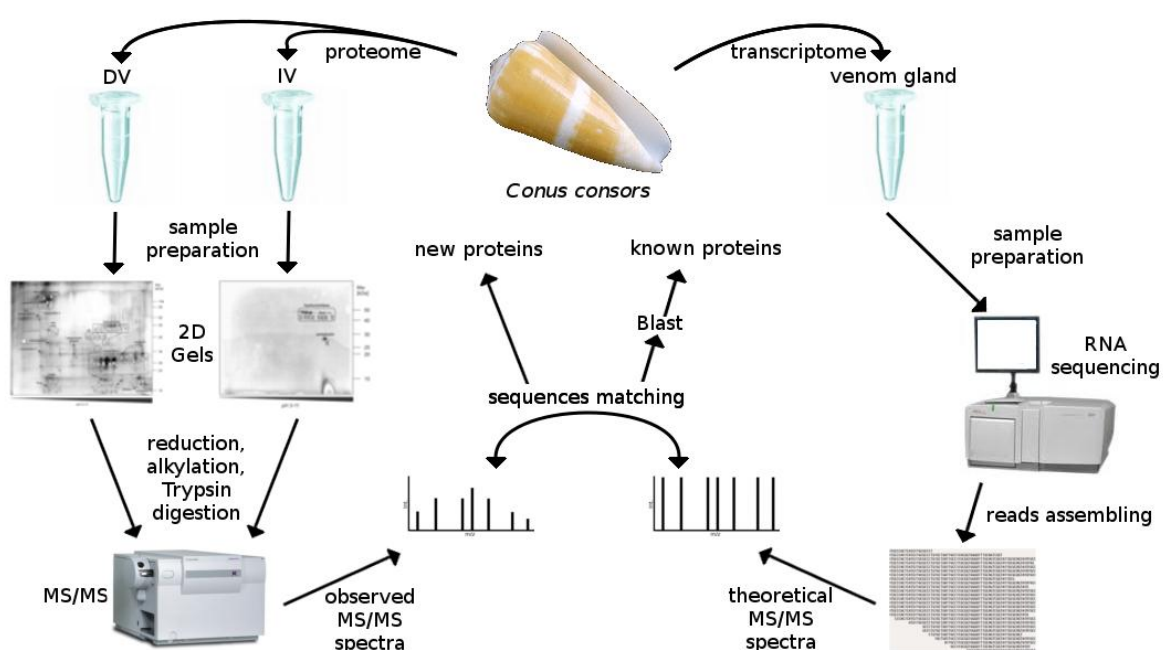
Personal contribution

All the MS/MS mass matching to the transcriptome, part of the data analysis and manuscript proofreading.

Abstract

For some decades, cone snail venoms have been providing peptides, generally termed conopeptides, that exhibit a large diversity of pharmacological properties. However, little attention has been devoted to the high molecular mass (HMM) proteins in venoms of mollusks. In order to shed more light on cone snail venom HMM components, the proteins of dissected and injected venom of a fish-hunting cone snail, *Conus consors*, were extensively assessed. HMM venom proteins were separated by two-dimensional polyacrylamide gel electrophoresis and analyzed by mass spectrometry (MS). The MS data were interpreted using UniProt database, EST libraries from *C. consors* venom duct and salivary gland, and their genomic information.

Numerous protein families were discovered in the lumen of the venom duct and assigned a biological function, thus pointing to their potential role in venom production and maturation. Interestingly, the study also revealed original proteins defining new families of unknown function. Only two groups of HMM proteins passing the venom selection process, echotoxins and hyaluronidases, were clearly present in the injected venom. They are suggested to contribute to the envenomation process. This newly devised integrated HMM proteomic analysis is a big step toward identification of the protein arsenal used in a cone snail venom apparatus for venom production, maturation, and function.



1. Introduction

Most marine gastropods of the Conoidea superfamily (Neo-gastropoda) have developed a venom apparatus enabling prey capture, defense, and competition. Among more than 15,000 species currently estimated from the Neogastropoda clade, the unique genus *Conus* (more than 700 species) is an invaluable resource of new bioactive compounds. Amid the hundreds of components contained in one venom, linear or disulfide-rich conopeptides have been particularly attractive on account of their exquisite action on a wide variety of ion channels and receptors (Olivera, 2006). Conopeptides are low molecular mass (LMM) compounds frequently exhibiting post-translational modifications such as C-terminal amidation, hydroxylation, carboxylation, bromination, epimerization, and glycosylation (Buczek, Bulaj, & Olivera, 2005; Marx, Daly, & Craik, 2006). The best known are the α -, μ - and ω -conotoxins that act specifically on nicotinic acetylcholine receptors and sodium and calcium ion channels, respectively. Several conotoxins have already demonstrated therapeutic potential in preclinical and clinical trials; one of them, ω -MVIIA (Ziconotide) from *C. magus*, has been approved for chronic pain treatment and commercialized (Prialt) (Miljanich, 2004).

Compared to the wealth of information about medically interesting conopeptides, the high molecular mass (HMM) components of cone snail venoms are still poorly characterized. Nevertheless, the first reports about these proteins appeared some decades ago when proteolytic, acetylcholinesterase, and phosphodiesterase activities were detected in cone snail venoms (Marsh, 1971; Balbin, 1980; Miranda, 1982). In cone snails, the venom is synthesized and stored in a long venom duct, also called the venom gland. The salivary gland and the epithelial zone between the venom duct and the pharynx have been suggested to contribute to the production of some components found in the injected venom (Biggs, Olivera, & Kantor, 2008; Biass et al., 2009). Following secretion into the lumen of the venom

duct, usually in the form of granules (Figure 36), venom components may undergo further in situ maturation by luminal enzymes. The viscous fluid squeezed out of the excised venom duct is named here the dissected venom (DV). By a currently unknown mechanism, the DV is “filtered” in such a way that only a selected cocktail of molecules is finally injected into a prey or a predator. Such “filtered” venom is termed injected venom (IV).

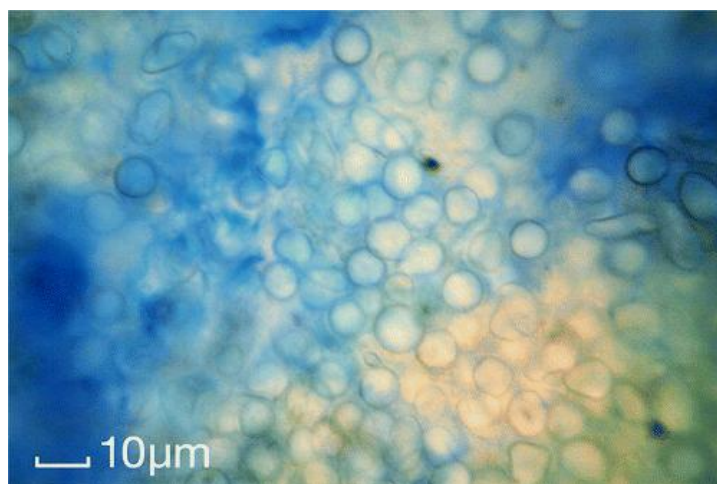


Figure 36: Venom components are partially or completely secreted into the lumen of the venom duct in the form of granules.

After deposition of a small drop of venom from the venom gland, granules were observed under optical microscopy at a 1000-fold magnification and with Coomassie blue coloration.

Many of the so far characterized HMM *conus* venom gland proteins have been classified as intracellular proteins. Tentatively, they have been proposed to originate from cells of the venom apparatus ruptured during preparation of the venom extract. An example of such a protein is a γ -glutamyl carboxylase (Stanley, Stafford, Olivera, & Bandyopadhyay, 1997; Czerwicz et al., 2002) that catalyzes post-translational carboxylation of specific glutamate residues in some conopeptides to γ -carboxyglutamate residues.

It is not clear whether the CRISP-related Tex31 protease is a DV-residing protein or not. This enzyme was isolated from a whole venom duct homogenate of *C. textile* (Milne, Abbenante,

Tyndall, Halliday, & Lewis, 2003) and suggested to be a specific protease acting on the pro-peptide cleavage site of conotoxin TxVIA. Two homologues of Tex31, Mr30 and GlaCRISP, were isolated from the crude venom of *C. marmoreus*; however, no proteolytic activity could be demonstrated using a Tex31 substrate (Hansson, Thamlitz, Furie, Furie, & Stenflo, 2006; Qian, Guo, & Chi, 2008). It is possible though that Tex31 orthologs in *C. marmoreus* possess substrate specificities that differ from those of Tex31. Tex31 homologues have also been detected recently in the *C. victoriae* venom gland (Safavi-Hemami et al., 2011). In this extensive proteomic survey hundreds of proteins were identified in the whole venom apparatuses of *C. novaehollandiae* and *C. victoriae*, but this approach did not allow inference of which proteins truly originate from DV.

In contrast, a protein catalyzing the hydrolysis of an ester bond at the sn-2 position in phospholipids is definitely part of the HMM protein complement of the DV. Conodipine-M, as this enzyme was named, has been isolated and characterized from *C. magus* venom squeezed from dissected venom ducts (McIntosh et al., 1995). Conodipine-M belongs to the family of secreted phospholipases A2 (sPLA2) and defines the structurally unique group IX of these enzymes (Schaloske & Dennis, 2006).

Apart these few cases, practically nothing else is known about the HMM proteins in the injected venom. To obtain further insights into the enzyme machinery involved in the venom maturation and envenomation process, we have focused on characterization of the HMM components in the DV and the IV of a piscivorous cone snail *C. consors*. Interpretation of the MS results, supported by extensive transcriptomic and genomic data accumulated from the same species, revealed the presence of proteins belonging to several families in the DV. In contrast, only two groups of proteins were found in the IV, and these are likely to be a part of the venom arsenal. While proteins secreted into the lumen of the venom duct are proposed to be mainly involved in further maturation of the pharmacologically active conopeptides, some HMM proteins are suggested to participate in selection of the components for the IV or

to be a part of the venom arsenal itself. Some proteins identified in this study define new protein families.

2. Materials and Methods

2.1 Reagents

Acetonitrile (ACN, from Fisher Scientific Ltd., Loughborough, UK), trifluoroacetic acid (TFA, from Pierce–Perbio, Lausanne, Switzerland), and formic acid (Acros Organics, Geel, Belgium) were of HPLC gradient grade or higher. Deionized water was purified using a Milli-Q system (Millipore Corp., Billerica, MA, USA). When needed, each solvent prepared was filtered and sonicated before use.

2.2 Specimen Collection

The specimens of *C. consors* were collected in the Chesterfield Islands (New Caledonia) during the CONFIELD expedition in June 2007. They were shipped to Atheris Laboratories (Geneva, Switzerland) where they were kept in an aquarium.

2.3 Dissected Venom

Venom from a batch of *C. consors* was obtained by dissection on ice of the venom duct from 19 specimens. Contents of venom ducts were squeezed out, pooled and diluted in 500 μ L of 10% (v/v) acetic acid, lyophilized, and stored at -80°C until use. Total protein content was estimated at ~ 7 mg per batch. Venom samples dedicated for the two-dimensional polyacrylamide gel electrophoresis (2DGE) analysis were not desalted. The lyophilized venom sample was not completely soluble in distilled water. The complete dissected venom sample is referred to hereafter as the raw DV and its water-soluble part as the DV.

2.4 Injected Venom

Live *C. consors* specimens were kept in two separate aquaria. A batch of *C. consors* IV was prepared using 6 cone snails. The venom was obtained using a milking procedure adapted from Hopkins and colleagues (Hopkins et al., 1995). A few microliters of venom were obtained, and the operation was repeated 67 times to provide a single venom pool. A single specimen was milked once or twice a week. As soon as pools of IV were obtained, these were centrifuged, lyophilized, and stored at -80°C . The protein content in an average pool was 400–500 μg .

2.5 Two-Dimensional Polyacrylamide Gel Electrophoresis

Samples of DV were separated by 2DGE gel, first on the analytical level and then also on a larger scale. Analyses were focused on venom components exceeding a molecular mass of 10000 Da to complement the LC-MS/MS proteomic analysis of the same venom (Biass et al., 2009; Violette et al., 2012a).

2.6 Analytical 2DGE

Analytical 2DGE was used to optimize separation conditions of *C. consors* venom by the Taguchi approach (Khoudoli, Porter, Blow, & Swedlow, 2004). In the first dimension, molecules were separated by isoelectric focusing (IEF) according to their isoelectric points (pI). The 7 cm immobilized pH gradient (IPG) strips, covering pH ranges of 3–11 and 4–7 (GE Healthcare, Amersham Biosciences), were rehydrated overnight with a sample in optimized rehydration buffer (ORB) containing 7 M urea, 2 M thiourea, 30 mM Tris, 0.75% (or 2.0% (v/v) in the case of 4–7 and 6–11 IPG strips), ampholytes, 0.25% (w/v) ASB-14, 3.0% (w/v) CHAPS, 0.002% (w/v) bromphenol blue and 12 $\mu\text{L/mL}$ DeStreak Reagent (GE Healthcare, Amersham Biosciences). The 7 cm IPG strip pH 6–11 (GE Healthcare,

Amersham Biosciences) was rehydrated overnight in ORB and the sample (8–10 µg) applied via cup loading. IEF was performed on Ettan IPGphor II (GE Healthcare, Amersham Biosciences) for a total of 6000 Vhrs (3–11NL IPG strip) or 8000 Vhrs (4–7 or 6–11 IPG strips). The current was restricted to 50 mA per strip. Following the IEF, the focused IPG strips were incubated for 15 min with reducing buffer containing 65 mM DTT, 75 mM Tris-HCl, 6 M urea, 4% (w/v) SDS, 30% (v/v) glycerol and 0.002% (w/v) bromphenol blue. Reduced proteins were then alkylated for 15 min with 135 mM iodoacetamide in the same buffer, but without DTT. The second dimension SDS PAGE was performed on 12.5% (w/v) polyacrylamide (PA) gels cast according to Laemmli (Laemmli, 1970) or 10% (w/v) PA gels cast in a gel buffer containing 150 g Tris in 0.6 M HCl,²² and run in a Tris/ taurine buffer system at 10 mA/gel as described (Tastet et al., 2003). Protein spots were visualized by silver staining.

2.7 Semi-preparative 2DGE

Following optimization of separation parameters on the analytical level, samples of *C. consors* venom were analyzed on a larger scale for subsequent MS analysis. Lyophilized DV was dissolved in rehydration buffer. A 24 cm IPG strip 3–11 NL was passively rehydrated overnight with 350 µg sample in 450 µL of ORB. IEF was carried out on an Ettan IPGphor II under the following conditions: 150 V for 3 h, 300 V for 3 h, 300–1000 V gradient for 6 h, 1000–10000 V gradient for 1 h and 10000 V for 3 h, a total of 40000 Vhr. The current was restricted to 75 µA/strip. The focused proteins were reduced, alkylated with iodoacetamide and analyzed in the second dimension SDS-PAGE on a 10% (w/v) PA gel. SDS PAGE was carried out in a Tris/taurine system using the Ettan DALT six unit (GE Healthcare, Amersham Biosciences) at 15 °C, applying 2 W/gel for 1 h and 17 W/gel for a further 5 h. Proteins in the gel were visualized by imidazole-SDS-Zn²⁺ reverse staining (Castellanos-Serra, Vallin, Proenza, Le Caer, & Rossier, 2001). This detection technique is similar in sensitivity to MS-

compatible silver staining. The stained gel was scanned by Image Scanner using LabScan 5 software (GE Healthcare, Amersham Biosciences) and the scan analyzed by Image Master 2D Platinum 6.0 software (GE Healthcare, Amersham Biosciences). The spots were excised from the gel automatically using Ettan Spot Picker (GE Healthcare, Amersham Biosciences) and stored at -20°C before further processing.

Due to low spot-pattern complexity, the IV samples for MS analysis were prepared using 7 cm IPG strips as described above, but using more material (20 μg).

2.8 In-Gel Digestion and Identification of Protein Digest Products

Immediately prior to MS analysis, gel pieces containing protein spots were thawed and destained in 70% (v/v) Tris/Gly (50 mM/0.3 M) and 30% (v/v) ACN. The gel pieces were then washed consecutively with 10 mM NH_4HCO_3 , 10 mM $\text{NH}_4\text{HCO}_3/\text{ACN}$ (1:1/v:v) and 100% ACN, then vacuum-dried. Proteins were digested in gel with mass spectrometry grade modified trypsin (Promega) in 25 mM NH_4HCO_3 at 37°C overnight. Resulting peptides were extracted with 50% (v/v) ACN/5% (v/v) formic acid, concentrated in vacuum to 15 μL and stored at -20°C . Extracts were analyzed using an electrospray ionization (ESI) ion trap-MS (MSD Trap XCT Plus, Agilent).

2.9 ESI-Ion Trap-MS analysis

LC-MS/MS analyses were performed on an ion trap mass spectrometer 1200 series HPLC-Chip-LC/MSD Trap XCT Ultra (Agilent Technologies, Waldbronn, Germany) with an ESI source operating in positive mode, controlled by ChemStation LC 3D systems Rev. B.01.03 SR1 (Agilent Technologies, Santa Clara, CA, USA) and LC/MSD Trap Control software version 6.0 (Bruker Daltonik GmbH, Bremen, Germany). An HPLC-Chip (Agilent Technologies, Waldbronn, Germany), comprising a 40 nL enrichment column and a 43 mm x

75 μm analytical column packed with Zorbax 300SB-C18 5 μm particles, was used for separating tryptic mixtures. Peptides were loaded onto the enrichment column in 97% (v/v) solvent A (0.1% (v/v) formic acid in water) and 3% (v/v) B (0.1% (v/v) formic acid in ACN) at 4 $\mu\text{L}/\text{min}$, and eluted from the analytical column with a gradient from 3% (v/v) B to 50% (v/v) B in 41 min, followed by a steep gradient to 90% (v/v) B in 1 min at a flow rate of 0.35 $\mu\text{L}/\text{min}$. MS acquisitions were carried out from 400 to 2200 m/z , followed by MS/MS scans of the five most abundant ions in each MS scan.

2.10 Bioinformatics/Data Mining

LC-ESI-MS spectral data, obtained as Mascot generic files (mgf), were searched against the *C. consors* transcriptome data file using Phenyx v.2.6.2 software (Genebio, Geneva, Switzerland). This operation was performed with the following parameters: parent ion error tolerance of 0.5 Da, maximum enzyme miscleavage of 2, minimum peptide length of 6, minimum peptide z-score of 7, possible proline hydroxylation and methionine oxidation PTMs (in variable positions) and carboxamidomethylcysteine PTM for all cysteine residues. All established *C. consors* amino acid sequences were then searched against the UniProt protein database to identify HMM proteins by homology. UniProt Gene Ontology (GO) annotations of homologous proteins were used to classify identified *C. consors* proteins into functional categories, using the 'GO-MIPS funcat conversion table' (<ftp://ftp.geneontology.org/pub/go/external2go/mips2go>).

To find the orthologs of *C. consors* HMM and LMM genes and the potential orthologs of the novel conoproteins, the non-redundant (NR), EST, GSS, HTGS, TSA, ENV_NT, WGS and genome databases of the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>), as well as the Ensembl (<http://www.ensembl.org>) and DOE Joint Genome Institute databases (<http://www.jgi.doe.gov>) were searched. Locally installed *C. consors* genome and diverse transcriptome (venom duct- and salivary gland-specific)

databases were also searched for the HMM and LMM proteins. Comparisons were performed using different BLAST tools (Gertz, Yu, Agarwala, Schaffer, & Altschul, 2006) with the E-value cut off set to 10^{-5} , and other parameters to default settings. DNA sequences were translated using the Translate program (<http://www.expasy.org/tools/dna.html>). The presence of signal peptides was verified using the SignalP 4.0 Server (Petersen, Brunak, von, & Nielsen, 2011) at <http://www.cbs.dtu.dk/services/SignalP/>. The conserved protein domains in the newly discovered HMM and LMM representatives were identified using the SMART (smart.embl-heidelberg.de), InterPro (<http://www.ebi.ac.uk/interpro/>) and Pfam (pfam.janelia.org) domain databases. The protein sequences were aligned using Clustal W2 (Larkin et al., 2007).

3. Results and Discussion

3.1 Analysis of Cone Snail Venoms by 2DGE

Venoms, either DV or IV, of *C. consors* were available in limited amounts. Conditions to achieve optimal 2DGE separation were therefore initially sought using analytical scale 2DGE (7 × 7 cm gel) and the DV of a less rare worm-hunting cone snail *C. quercinus*. Using the Taguchi approach (Khoudoli et al., 2004), ampholyte concentration, zwitterionic detergent concentrations (CHAPS and ASB-14) and pH range parameters were varied in the first, IEF dimension. In the second, SDS PAGE dimension, buffer systems, Tris/glycine or Tris/taurine, and acrylamide concentration were varied. The resolution was much better if the venom was not desalted prior to analysis. Attaining conditions to resolve *C. quercinus* DV into the maximal number of discrete spots (see methods section), the raw *C. consors* DV was also analyzed under the same conditions. Interestingly, the 2DGE HMM protein pattern of the worm-hunting cone snail *C. quercinus* was much more complex than that of the fish-hunting cone snail *C. consors*. Using silver staining for detection, the *C. quercinus* DV was resolved

into 129 discrete spots and *C. consors* DV into 92 (Figure 37). Substantial variation in DV complexity between the worm-hunting and the fish-hunting cone snail reflects specific evolutionary traits of their venom apparatus.

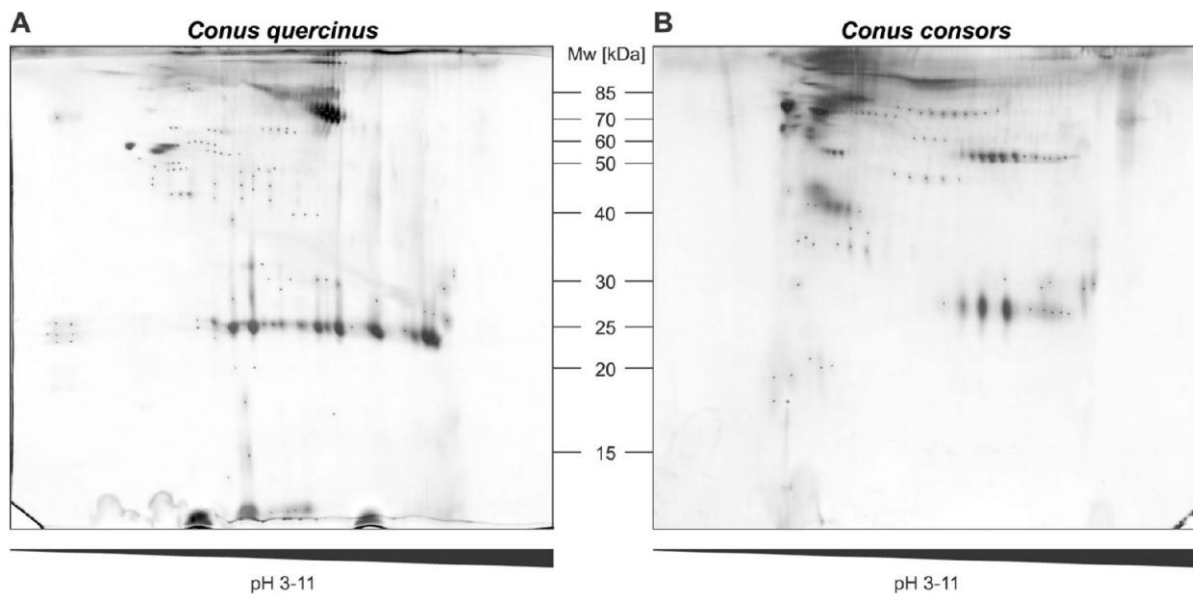


Figure 37: Venom composition of the fish-hunting cone snail *C. consors* is less complex than that of the worm-hunting cone snail *C. quercinus*.

Analytical two-dimensional gels of the DVs from *C. quercinus* (A) and *C. consors* (B) are shown. IEF of the IV or the DV was performed under denaturing conditions using a 3–11 NL IPG strip followed by 10% (w/v) SDS-PAGE in Tris/Taurin buffer in a perpendicular dimension. As revealed by silver staining, the vermivorous *C. quercinus* contains many more HMM proteins in its DV (129) than the piscivorous *C. consors* (92).

As expected, *C. consors* IV contained far fewer HMM protein components (only 9) than the DV. This suggests that most of the HMM protein venom components are not involved directly in the envenomation process but rather participate in venom maturation or some other function.

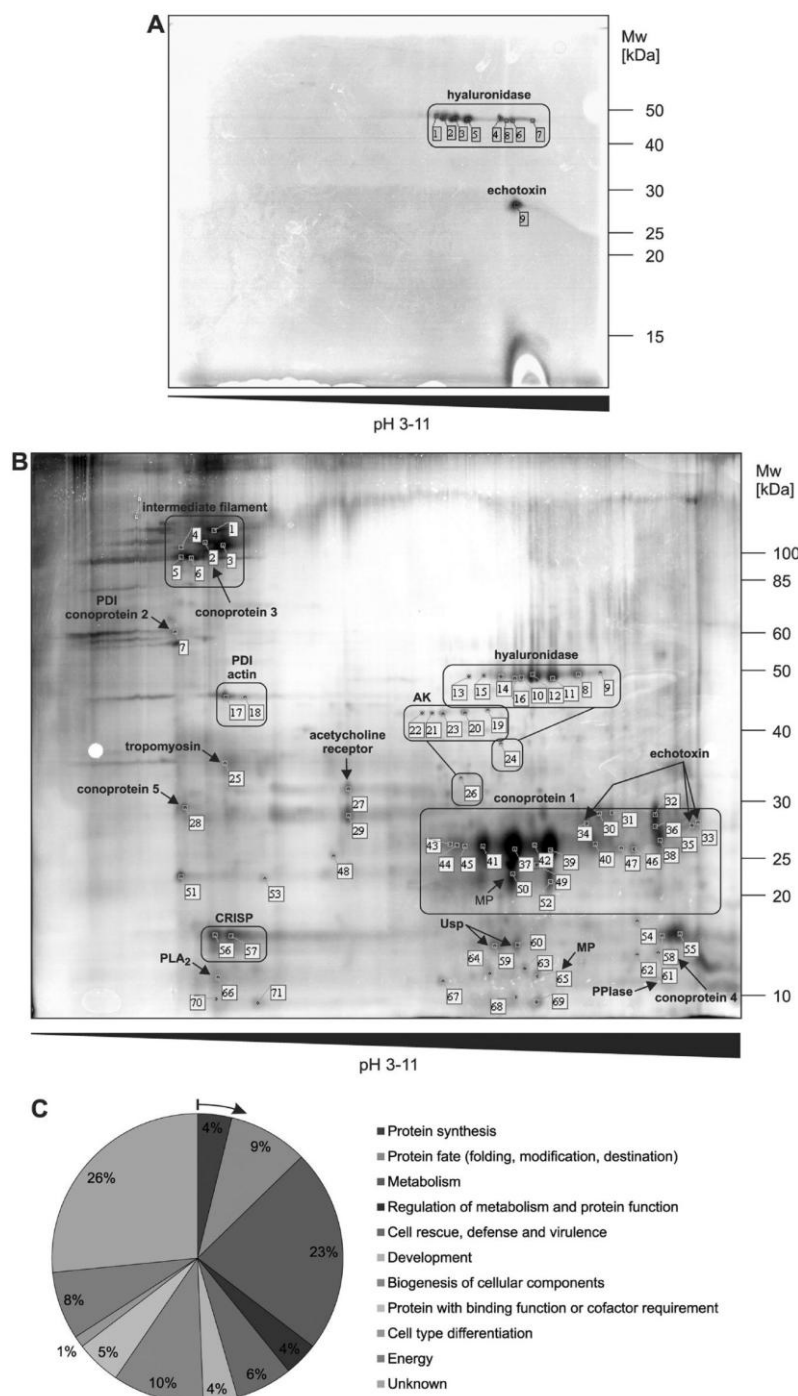


Figure 38: Exploring the proteomes of *C. consors* IV and DV

by two-dimensional gel electrophoresis. IEF of the IV or the DV was performed under denaturing conditions using a 3–11 NL IPG strip followed by 10% (w/v) SDS-PAGE in Tris/Taurin buffer in the perpendicular dimension. Gels were reverse stained using imidazole-SDS-Zn²⁺ staining protocol. Nine spots were detected in the IV (**A**) and 71 in the DV (**B**). Spots were excised, reduced, alkylated, digested by trypsin, and analyzed by MS/MS. Results of mass spectrometry data processing are collected in Tables 1 and 2 and in the Supplementary File. Numbers of the spots on gels correspond to spot numbers in the different Tables and Supplementary File. Proteins identified in the *C. consors* DV were grouped into 11 categories according to their UniProt GO annotation (**C**).

The same separation conditions were then used on a larger, semi-preparative scale. When running the 2DGE analysis with about ~40-fold more of the raw DV, smearing appeared and resolution was largely lost. This effect was related to the insoluble fraction of the raw material, since the problem was avoided by analyzing the water-soluble fraction only. No difference between spot patterns of the raw DV and its water-soluble fraction was observed on the analytical scale. The 71 well-defined spots obtained by imidazole-SDS-Zn²⁺ reverse staining were further processed for MS identification (Figure 38B). Described negative effect was not so apparent in the case of the IV: the spot pattern obtained was similar to the analytical one, but only 9 clearly defined spots were visible. All of them were prepared for MS analysis (Figure 38A).

3.2 HMM Proteins in *C. consors* IV

The 2DGE spot pattern of the IV was much simpler than that of the DV. This could be explained by most of the proteins being located in the venom granules, and/or by a selection process excluding most of the HMM protein components from the venom prior to injection. From all 9 spots, useful sequence information was obtained (Figure 38A and Table 13), but surprisingly only two protein families were found: hyaluronidases and echotoxin-like proteins.

Eight of the spots were identified as representing hyaluronidase isoforms (EC 3.2.1.35). A more detailed analysis by LC-ESI-MS/MS of the IV revealed the presence of 14 isoforms of this enzyme, with two alternatively glycosylated paralogs (Violette et al., 2012b). Hyaluronidases degrade hyaluronan (also called hyaluronic acid or hyaluronate), a high molecular mass polysaccharide abundantly present in the extracellular matrix of most connective tissue (Girish & Kemparaju, 2007). Hyaluronidases are not toxic per se but probably help to spread injected venom in the tissue of a prey or a predator (Tu & Hendon, 1983; Kemparaju K., Girish K.S., & Nagaraju S., 2010).

Table 13: HMM proteins identified in the *C. consors* IV

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
1	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-52}	8	91.7	Glycosyl hydrolase 56
2	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-52}	5	55.3	Glycosyl hydrolase 56
3	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-52}	16	168.6	Glycosyl hydrolase 56
4	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	13	143.4	Glycosyl hydrolase 56
5	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	16	168.0	Glycosyl hydrolase 56
6	Hyaluronidase PH-20 [P23613]	<i>Cavia porcellus</i>	3.00×10^{-30}	5	48.1	Glycosyl hydrolase 56
7	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	10	101.3	Glycosyl hydrolase 56
8	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	11	113.6	Glycosyl hydrolase 56
9	Echotoxin-2 [Q76CA2]	<i>Cymatium echo</i>	1.00×10^{-23}	7	77.7	Sea anemone actinoporin

The ninth spot of the IV gel corresponded to an echotoxin-like protein, a family also found in the DV in two isoforms. Echotoxins were originally purified from the salivary gland of the marine gastropod *Monoplex echo* (Shiomi, Kawashima, Mizukami, & Nagashima, 2002). These proteins were demonstrated to be hemolytic and lethal to mice. They are structurally homologous to actinoporins, pore-forming hemolysins from sea anemones, and appear to act in the same way (Kawashima, Nagai, Ishida, Nagashima, & Shiomi, 2003; Gunji, Ishizaki, & Shiomi, 2010).

3.3 HMM Proteins in the *C. consors* DV

ESI-ion trap MS analysis of 54 out of the total of 71 DV spots gave useful sequence information (Figure 38B and Table 14). As shown in Table 14, some spots contained more than one protein. Spots from which very poor or no MS spectra were obtained might simply

contain too low an amount of protein or might result from staining of molecules other than proteins. Indeed, zinc-imidazole reverse staining can detect other biomolecules besides proteins on gels, for example nucleic acids or lipopolysaccharides. In addition, other artefacts of a nonbiomolecular character have been described for this staining protocol (Castellanos-Serra et al., 2001). Out of 54 spots, partial sequences of 77 different proteins were acquired. 56 of these proteins were identified by searching against *C. consors* venom duct EST library and the UniProt database, leaving 21 unassigned. Sequences belonging to the same protein were frequently found in spots differing in either pI or molecular mass position on the gel. Different pI values certainly indicate the existence of more isoforms of a particular protein in the venom, while different molecular masses may also reflect proteolytic trimming or degradation of a protein. Taking isoforms into account, 23 different proteins, belonging to 22 different protein families, were identified in the *C. consors* (Table 15).

Table 14: HMM proteins identified in the *C. consors* DV

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
1	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	3.00×10^{-32}	6	66.5	Intermediate filament
2	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	3.00×10^{-32}	5	54.8	Intermediate filament
	Conoprotein 3	Unknown		2	13.4	Unknown
3	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	5.00×10^{-46}	5	43.4	Intermediate filament
	Actin [Q964E0]	<i>Biomphalaria tenagophila</i>	3.00×10^{-159}	2	20.1	Actin
4	Non-neuronal cytoplasmic intermediate filament protein	<i>Helix aspersa</i>	3.00×10^{-27}	2	21.8	Intermediate filament

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
	[P22488]					
5	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	3.00×10^{-32}	6	66.1	Intermediate filament
	Protein sidekick [O97394]	<i>Drosophila melanogaster</i>	4.00×10^{-4}	2	19.3	Protein sidekick (immunoglobulin superfamily)
6	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	3.00×10^{-32}	7	77.8	Intermediate filament
	51 kDa retrograde protein [Q6QUW1]	<i>Lymnaea stagnalis</i>	4.00×10^{-17}	3	21.2	Intermediate filament
7	PDI [P05307]	<i>Bos taurus</i>	1.00×10^{-159}	30	318.9	PDI
	Collagen alpha-5(VI) chain [A6H584]	<i>Mus musculus</i>	1.00×10^{-29}	3	50.1	Type VI collagen
	Non-neuronal cytoplasmic intermediate filament protein [P22488]	<i>Helix aspersa</i>	3.00×10^{-27}	2	18.0	Intermediate filament
	Conoprotein 2	Unknown		2	14.7	Unknown
8	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	8.00×10^{-50}	4	52.7	Glycosyl hydrolase 56
9	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	13	139.7	Glycosyl hydrolase 56
10	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	8.00×10^{-50}	7	88.5	Glycosyl hydrolase 56
11	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	17	179.9	Glycosyl hydrolase 56
12	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	18	193.1	Glycosyl hydrolase 56
13	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	10	108.0	Glycosyl hydrolase 56
14	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-52}	18	198.4	Glycosyl hydrolase 56
	Elongation factor 1-gamma [Q90YC0]	<i>Carassius auratus</i>	6.00×10^{-124}	3	37.8	Elongation factor

6. Peptidomics and Proteomics

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
15	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	16	171.9	Glycosyl hydrolase 56
16	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-52}	15	190.8	Glycosyl hydrolase 56
17	Actin [Q964E0]	<i>Biomphalaria tenagophila</i>	∞	20	202.1	Actin
	PDI [P05307]	<i>Bos taurus</i>	1.00×10^{-159}	2	17.3	PDI
18	Actin [Q964E0]	<i>Biomphalaria tenagophila</i>	∞	17	186.3	Actin
	PDI [Q8R4U2]	<i>Cricetulus griseus</i>	1.00×10^{-150}	3	30.4	PDI
19	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	12	125.4	ATP:guanidino phosphotransferase
20	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	29	330.5	ATP:guanidino phosphotransferase
21	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	10	112.7	ATP:guanidino phosphotransferase
22	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	7	75.3	ATP:guanidino phosphotransferase
23	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	28	311.0	ATP:guanidino phosphotransferase
24	Hyaluronidase [Q6RHW4]	<i>Sus scrofa</i>	3.00×10^{-53}	2	23.8	Glycosyl hydrolase 56
25	Tropomyosin [P42636]	<i>Biomphalaria glabrata</i>	6.00×10^{-19}	2	17.1	Tropomyosin
26	AK [O15990]	<i>Liolophura japonica</i>	1.00×10^{-133}	6	75.9	ATP:guanido phosphotransferase
27	Neuronal acetylcholine receptor subunit alpha-7 [Q05941]	<i>Rattus norvegicus</i>	1.00×10^{-15}	2	43.6	Ligand-gated ion channel
28	Conoprotein 5	Unknown		2	21.1	Unknown

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
30	Conoprotein 1	Unknown		5	73.7	Unknown
31	Conoprotein 1	Unknown		4	54.6	Unknown
32	Conoprotein 1	Unknown		2	30.7	Unknown
	Succinate dehydrogenase [Q9YHT2]	<i>Gallus gallus</i>	3.00×10^{-72}	2	23.8	Succinate dehydrogenase/fumara te reductase iron–sulfur protein
33	Echotoxin-2 [Q76CA2]	<i>Cymatium echo</i>	1.00×10^{-23}	7	74.6	Sea anemone actinoporin
	Conoprotein 1	Unknown	-	2	24.3	Unknown
	60S ribosomal protein [Q6PC69]	<i>Danio rerio</i>	6.00×10^{-29}	2	19.9	Ribosomal protein L1P
34	Echotoxin-2 [Q76CA2]	<i>Cymatium echo</i>	6.00×10^{-13}	3	37.6	Sea anemone actinoporin
	Conoprotein 1	Unknown		2	26.4	Unknown
35	Echotoxin-2 [Q76CA2]	<i>Cymatium echo</i>	3.00×10^{-24}	5	64.9	Sea anemone actinoporin
	Conoprotein 1	Unknown		2	20.9	Unknown
36	Conoprotein 1	Unknown		3	36.1	Unknown
37	Conoprotein 1	Unknown		5	75.0	Unknown
38	Conoprotein 1	Unknown		5	66.1	Unknown
39	Conoprotein 1	Unknown		5	70.0	Unknown
	Glutathione S-transferase [Q03013]	<i>Homo sapiens</i>	2.00×10^{-11}	2	18.7	Glutathione S-transferase mu
40	Hyaluronidase-1 [Q12794]	<i>Homo sapiens</i>	5.00×10^{-37}	4	45.2	Glycosyl hydrolase 56
	Conoprotein 1	Unknown	-	2	25.2	Unknown

6. Peptidomics and Proteomics

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
	Glutathione S-transferase [Q9TSM5]	<i>Macaca fascicularis</i>	2.00×10^{-32}	2	18.9	Glutathione S-transferase mu
41	Conoprotein 1	Unknown		5	71.9	Unknown
42	Conoprotein 1	Unknown		5	73.4	Unknown
43	Transmembrane emp24 domain-containing protein [Q295B2]	<i>Drosophila pseudoobscura pseudoobscura</i>	2.00×10^{-40}	5	54.3	EMP24/GP25L
	Conoprotein 1	Unknown		3	37.2	Unknown
44	Conoprotein 1	Unknown		2	24.8	Unknown
	Hyaluronidase [Q62803]	<i>Rattus norvegicus</i>	4.00×10^{-20}	2	23.4	Glycosyl hydrolase 56
50	Zinc-metalloproteinase [P42674]	<i>Paracentrotus lividus</i>	7.00×10^{-30}	10	110.8	Peptidase M12A
	Zinc metalloproteinase [Q20176]	<i>Caenorhabditis elegans</i>	1.00×10^{-20}	6	72.5	Peptidase M12A
	Conoprotein 1	Unknown		5	69.6	Unknown
52	Conoprotein 1	Unknown		5	71.9	Unknown
56	Peptidase inhibitor 16 [Q58D34]	<i>Bos taurus</i>	4.00×10^{-32}	4	45.7	CRISP
57	Peptidase inhibitor 16 [Q58D34]	<i>Bos taurus</i>	1.00×10^{-22}	3	47.0	CRISP
58	Conoprotein 4	Unknown		2	25.9	Unknown
59	Universal stress protein [Q57951]	<i>Methanocaldococcus jannaschii</i>	2.00×10^{-7}	4	43.9	Universal stress protein A
	Transgelin-2 [Q9WVA4]	<i>Mus musculus</i>	9.00×10^{-7}	5	42.4	Calponin
	40S ribosomal protein [P62844]	<i>Sus scrofa</i>	5.00×10^{-40}	2	12.2	Ribosomal protein S19P

spot no.	homologous protein	species	Blast <i>E</i> -value	matched peptides	best EST score	protein family
60	Universal stress protein [Q57951]	<i>Methanocaldococcus jannaschii</i>	1.00×10^{-7}	4	43.3	Universal stress protein A
61	PPI [P54985]	<i>Blattella germanica</i>	5.00×10^{-73}	4	29.2	Cyclophilin-type PPIase
65	Zinc metalloproteinase [Q20176]	<i>Caenorhabditis elegans</i>	1.00×10^{-16}	2	19.9	Peptidase M12A
66	Conodipine-M [Q9TWL8]	<i>Conus magus</i>	1.00×10^{-18}	4	42.8	sPLA ₂

3.3.1 Typical Intracellular Proteins.

Despite the mode of collecting venom (see Methods section), the 13 HMM proteins are structural homologues of proteins otherwise found exclusively inside cells (Table 15). Their extracellular location here is unexpected. They could come from the mechanical rupture of some cells lining the venom duct, thus being a kind of contamination. They could also arise from the disruption of the venom granules (Figure 36) during the preparation of dissected venom. It is also possible that some of these proteins are in fact secreted from cells, even though they do not possess any known export signals. An example of such a scenario could be the case of arginine kinases (AK). An AK-like protein has been described as a component of venom of the solitary spider wasp *Cyphononyx dorsalis* (Yamamoto, Arimoto, Kinumi, Oba, & Uemura, 2007). This protein was shown to have a paralytic activity against spiders, a natural prey of the wasp. Although the presence of AK in the *C. consors* IV was not confirmed, the possibility of a toxic function remains open. The *C. consors* DV contained multiple forms of AK. This enzyme was detected in 5 spots with the same molecular mass of 43 kDa but different pI values, ranging from 6.3 to 7.0 (spots 19–23 in Figure 38B and Table 14). One faint AK-containing spot was also found at 32 kDa (26 in Figure 38B and Table 14). This was probably due to proteolytic truncation of one of the 43 kDa AK forms. The complete

primary structure of *C. consors* AK has been deduced from its cDNA (Supplementary File). A well-known function of AK (EC 2.7.3.3) is the catalysis of the reversible transfer of the phosphate group from phospho-L-arginine to ADP to form ATP. Analysis of the *C. consors* genome revealed the presence of three AK genes. Two of them have gene structures typical of AKs, with 6 exons and 5 introns, while the third gene has completely lost the five introns and is intronless. Normal and intronless AK genes are very similar to the recently described *C. novaehollandiae* AK gene (Safavi-Hemami, Young, Williamson, & Purcell, 2010), while the more divergent AK gene is very similar to the *C. adamsonii* AK gene (unpublished data).

Table 15: HMM proteins of the *C. consors* venom listed by molecular function

function	protein	mass (kDa)	no. of spots
Protein synthesis			
Translation	40S ribosomal protein	15	1
	60S ribosomal protein	27.5	1
Translation	Elongation factor 1-gamma	48	1
Protein fate (folding, modification, destination)			
Protein folding	PDI	45	2
	PDI	60	1
Protein folding	PPI	14	1
Proteolysis	Zinc metalloproteinase	24	2
	Zinc metalloproteinase	12	1
Metabolism			
Carbohydrate lysis	Hyaluronidase	48	13
	Hyaluronidase	26	1
Lipid degradation	Conodipine-M	12	1

function	protein	mass (kDa)	no. of spots
Electron carrier	Succinate dehydrogenase	28	1
Transferase	Glutathione S-transferase	26	2
Regulation of metabolism and protein function			
Protease inhibitor	Peptidase inhibitor	16.5	2
Regulation of muscle contraction	Tropomyosin	34	1
Cell rescue, defense and virulence			
Pore-forming	Echotoxin-2	27.5	3
Response to stress	Universal stress protein	15	2
Development			
Cell adhesion	Protein sidekick	95	1
Developmental protein	Transmembrane emp24 domain-containing protein	26	1
Muscle development	Transgelin-2	15	1
Biogenesis of cellular components			
Structural molecule	Retrograde protein	95	1
Structural molecule	Non-neuronal cytoplasmic intermediate filament protein	120	1
	Non-neuronal cytoplasmic intermediate filament protein	106	3
	Non-neuronal cytoplasmic intermediate filament protein	95	2
	Non-neuronal cytoplasmic intermediate filament protein	60	1
Protein with binding function or cofactor requirement (structural or catalytic)			
Cell motility	Actin	45	2
	Actin	106	1

6. Peptidomics and Proteomics

function	protein	mass (kDa)	no. of spots
Neuronal receptor	Neuronal acetylcholine receptor subunit alpha-7	31	1
Cell type differentiation			
Cell adhesion	Collagen	60	1
Energy			
Catalytic activity	AK	43	5
	AK	32	1
Unknown			
	Conoprotein 1	22–28	16
	Conoprotein 1	48	1
	Conoprotein 2	60	1
	Conoprotein 3	16.5	1
	Conoprotein 4	29	1
	Conoprotein 5	108	1

3.3.2 Extracellular Proteins

The 10 HMM proteins identified as orthologs of extracellular proteins could be more important for the cone snail venom function. A homologue of the α -bungarotoxin-binding subunit of acetylcholine receptor (AChR), the $\alpha 7$ subunit, could be very interesting in this respect. It is known that α -conotoxins present in *C. consors* venom (Favreau et al., 1999) compete with α -bungarotoxin (a snake venom disulfide-rich polypeptide antagonist of AChR) for the same receptor binding site (Stiles, 1993). Binding of α -conotoxins to $\alpha 7$ AChR subunit-like protein in a cone snail venom duct can effectively modulate the concentration of α -conotoxins in the venom fluid, which may constitute a part of the venom component selection mechanism. Protein sidekick, a type I transmembrane protein that contains

immunoglobulin and fibronectin type III domains, and the type VI collagen that participates in cell adhesion and intracellular binding are, on the other hand, hardly likely to be involved directly in venom function.

3.3.3 Enzymatic Proteins

By homology to known enzymes, several proteins in the DV probably possess a number of enzymatic activities, such as proteinase, peptidylprolyl bond cis–trans isomerization, disulfide bond isomerization, phospholipase A2, and hyaluronan glucosaminidase activities. Some of these activities participate in venom maturation in the venom duct lumen and also directly in venom function. For example, depending on cone snail sensory inputs, proteases can activate or inactivate venom components. The activity of proteolytic enzymes can be fine-tuned by proteinase inhibitors discovered in the venom duct (spots 56 and 57 in Figure 38B and Table 14). Possibly regulated by external stimuli, the activities of peptidylprolyl cis–trans isomerase (PPI) and protein disulfide isomerase (PDI) can assist venom components to exchange between less and more active states, thus regulating venom function.

PPI (EC 5.2.1.8) catalyzes the cis–trans isomerization of proline imide peptide bonds in oligopeptides and accelerates the folding of proteins. This enzyme is important for proline-containing, cysteine-rich conotoxin biosynthesis (Safavi-Hemami, Bulaj, Olivera, Williamson, & Purcell, 2010). Major problems have been encountered in trying to fold conotoxins in vitro, very probably due to the specific environment and helper proteins involved in vivo (Safavi-Hemami et al., 2012). To advance procedures of folding of proline-containing cysteine-rich polypeptides, the complete primary structure of this enzyme was elucidated thanks to the transcriptome (Supplementary File). In this way, cloning and subsequent characterization of *C. consors* PPI as a folding catalyst was enabled.

In addition to PPI, PDI (EC 5.3.4.1) is another HMM protein of the cone snail venom that is extremely important for effective synthesis of cysteine-rich conotoxins (Bulaj et al., 2003).

Three forms of PDI were detected in the *C. consors* DV, one at 60 kDa and two at 45 kDa (spots 7, 17, and 18 in Figure 38B and Table 14). Three PDI genes were also found in the genome. Interestingly, Terrat and colleagues reported only two mRNAs encoding PDIs of about 45 kDa (Terrat et al., 2012). A similar multiform appearance of PDI was also reported in *C. novaehollandiae* and *C. victoriae* venoms (Safavi-Hemami et al., 2011).

PDI, possessing a specific endoplasmic reticulum (ER)-retaining motif, is usually regarded as an ER-residing protein. It has, however, also been found extracellularly (Mandel, Ryser, Ghani, Wu, & Peak, 1993). Given the mode of collecting DV used here, the detected forms of PDI are expected to be extracellular. Surface-associated PDI may catalyze disulfide bond exchange with conopeptides and may also, in this way, regulate the concentration of soluble conopeptides in the venom duct. Different isoforms of PDI may interact selectively with conopeptides, which introduces another level of selectivity in the regulation of conopeptide content in the injected venom.

A sequence homologous to that of conodipine M was found in spot 66 (Figure 38B and Table 14). Conodipine M is a structurally unique sPLA2 (EC 3.1.1.4) isolated from the dissected venom of *C. magus* (McIntosh et al., 1995). These enzymes have been recruited in venoms during the evolution of a number of venomous animals such as snakes, scorpions, spiders, and hymenoptera. Unexpectedly, conodipine M was detected uniquely in the dissected, but not the injected, venom. This result was in agreement with the lack and the presence of the phospholipase activity in the IV and the DV, respectively (T. Petan, personal communication). The role of the sPLA2 in the *C. consors* DV is not obvious at present but may not be involved directly in the envenomation process. It is important to note that venom processing and injection can vary with time, even for a single individual, in which dramatic variation has already been observed (Dutertre, Biass, Stocklin, & Favreau, 2010). Secreted PLA2 expression in the venom duct may also be a relic of ancient use in the IV, kept as a reservoir of bioactive material or used for prey digestion.

3.3.4 Proteins Recruited in the IV

Hyaluronidase sequences were detected in the DV in 9 discrete spots with about 50 kDa and pI values ranging from 6.7 to 8.3 (spots 8–16 in Figure 38B and Table 14). At the protein level, at least two hyaluronidase isoforms exist in the *C. consors* venom. Further, each of these is diversely glycosylated (Violette et al., 2012b). Four isomeric transcripts have been found at the transcriptomic level (Terrat et al., 2012). The molecular mass of the aglycone of the major hyaluronidase isoform was 48037.52 Da. Hyaluronidase forms detected in spots 24, 40, and 44 (Figure 38B and Table 14) were lighter. At 26 or 37 kDa, these proteins are probably degradation products of 50 kDa forms. Hyaluronidases hydrolyze hyaluronan, also reported as a constituent of some venoms (Alsop & Bettini, 1978). Due to its biophysical characteristics, hyaluronan raises the viscosity of the venom. Thus, if it is also present in the cone snail venom, then hyaluronidase activity in the DV may affect its viscosity and thus the efficacy of the venom injection.

The DV also contained three isoforms (spots 33, 34, and 35 in Figure 38B and Table 14) of a protein homologous to echotoxin (Shiomi et al., 2002; Kawashima et al., 2003; Gunji et al., 2010). At the transcriptomic level, the existence of three isoforms of echotoxin has been reported (Terrat et al., 2012). Present also in the IV and being toxic to mice, echotoxin-like proteins likely participate to the arsenal of *C. consors* molecules with a toxin function.

3.3.5 Novel Conoprotein Families

In 21 DV protein spots, partial amino acid sequences were matched against the transcriptomes or genome but could not be associated with any currently annotated protein families. These sequences are labelled as conoproteins 1–5 in Figure 38B and Table 14. Interestingly, judging from the gel staining intensity, conoprotein 1 appears to be the major component of the *C. consors* DV. Detected in 17 spots (Figure 38B and Table 14), it is represented by numerous isoforms in the DV. The fact that conoproteins 1–5 show no

homology with proteins present in the UniProt database may indicate that they are exclusive to mollusks. All five conoproteins have been found encoded in the *C. consors* genome. Due to the incomplete coverage of the *C. consors* genome sequence and the very short scaffolds, reconstruction of the full length transcripts of conoproteins 2, 3, and 5 was not possible. It was, however, possible for conoprotein 4. This consisted of 227 amino acids, including a signal peptide of 21 amino acids, as verified by SignalP 4.0 server. For conoprotein 1, a 203 amino acid long transcript was obtained that did not include the signal peptide. Conoproteins 1 and 4 are cysteine-rich proteins, the first possessing 13 and the second 15 cysteine residues.

Table 16: Distribution of Conoproteins 1–5 in genome and transcriptome databases

gene/protein name	<i>C. consors</i> genome	<i>C. consors</i> CC8 (VD + SG)	<i>C. consors</i> CC6 (VD + SG)	<i>C. consors</i> VD 6	<i>C. consors</i> SG 6	<i>C. adamsonii</i> (VD + SG)	NCBI NR	NCBI EST
Conoprotein 1	+	+ (1 hit)	+ (1 hit)	+ (1 hit)	0	0	0	0
Conoprotein 2	+	0	0	0	0	0	0	0
Conoprotein 3	+	+ (1 hit)	+ (1 hit)	+ (1 hit)	+ (1 hit)	0	0	0
Conoprotein 4	+	+ (3 hits)	+ (4 hits)	+ (1 hit)	+ (5 hits)	+ (2 hits)	0	+
Conoprotein 5	+	0	0	0	0	0	0	0

+: gene is present; 0: gene is absent; the frequency of gene in the transcriptome database is shown by the number of Blast hits. SG6: salivary gland transcriptome, VD6: venom duct transcriptome, CC6 and CC8: combined salivary gland and venom duct transcriptomes. These databases are deposited in the Conco database site (<http://conco.ebc.ee>)(Terrat et al., 2012).

The expression profiles of the conoprotein 1–5 genes showed that genes for conoproteins 3 and 4 were expressed in both the venom duct and the salivary gland, while the conoprotein 1 gene was expressed only in the venom duct. Transcripts of conoproteins 2 and 5 genes were not found in any of these tissues. In contrast to the large amount of conoprotein 1 detected in

the DV, the transcriptome data showed that conoproteins 1, 3, and 4 were expressed only at very low levels. Only 1–5 Blast hits were obtained for any of them (Table 16). Using the UniProt database for analysis, it appeared at first that these proteins are orphans, i.e., genes/proteins without any relatives. However, extending the search on the large NCBI EST and genome databases, conoprotein 4 was found to have numerous relatives, but only in mollusks (including gastropods, bivalves, and cephalopods). Conoprotein 4 could therefore

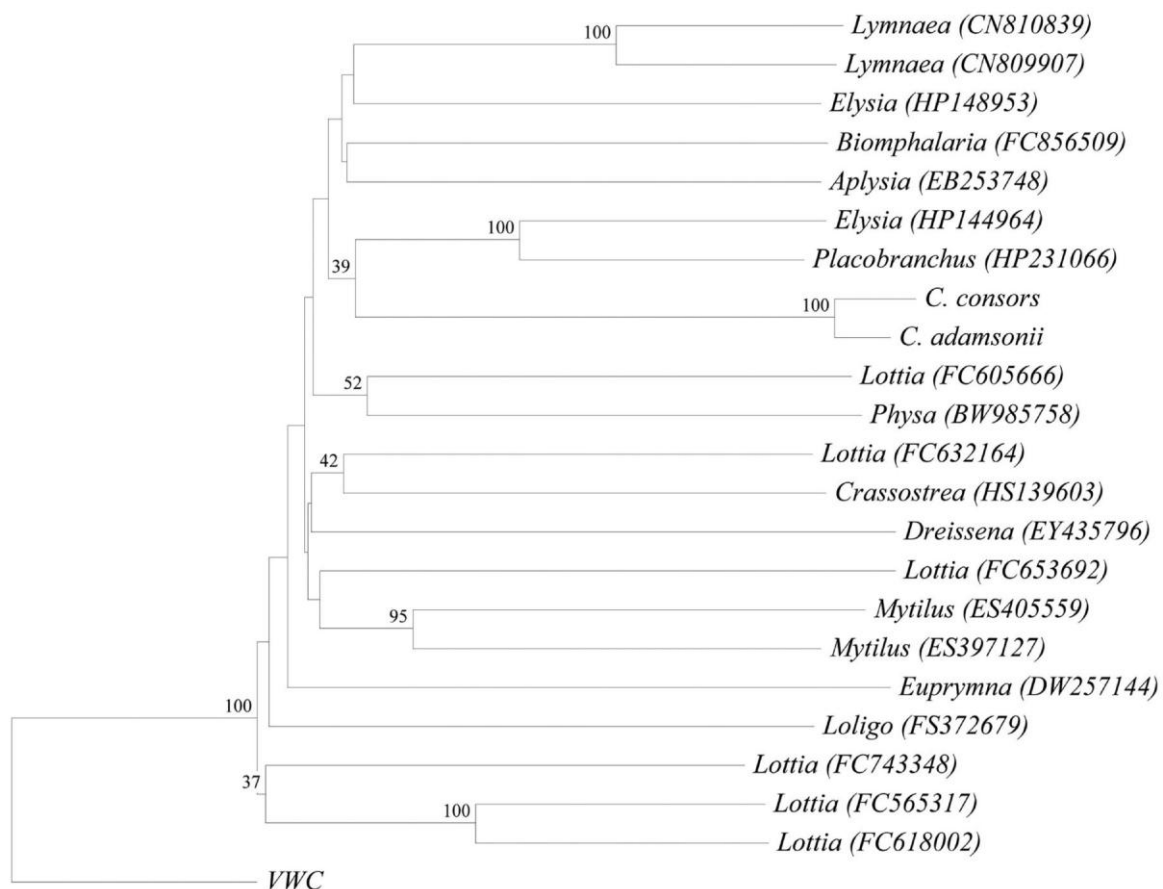


Figure 39: Evolutionary relationships within the mollusk-specific conoprotein 4 orthologous family.

The rooted neighbour-joining tree represents the bootstrap consensus following 1000 replicates; nodes with confidence values greater than 30% are indicated. Sequences were obtained from the EST database at the NCBI; genus names and accession numbers are included. Mammalian von Willebrand factor type C domain (VWC) was used as an outgroup.

be a lineage-specific gene that is mollusk-specific (Figure 39). The other four conoprotein genes are apparently orphan genes that are limited to the *C. consors* or some closely related piscivorous cone snail species. Conoprotein orphan genes may arise from duplication and rearrangement processes followed by rapid divergence (conoprotein 4) or by de novo evolution from non-coding genomic regions (conoproteins 1, 2, 3, and 5). This process appears to provide raw material continuously for the evolution of new gene functions that can become relevant for lineage-specific adaptations (Tautz & Domazet-Loso, 2011; Chang & Duda, Jr., 2012). Finally, these conoprotein families were revealed by our integrated proteomic approach, where MS data are matched against a set of nucleic acid sequences derived from the studied species. This approach is in marked contrast with previous studies where proteomic results are generated from protein databases related to other organisms (e.g., ref (Safavi-Hemami et al., 2011)).

3.4 Are *C. consors* HMM Venom Proteins Produced in the Venom Duct or in the Salivary Gland?

Analysis of the *C. consors* venom duct transcriptome revealed that mRNAs coding for some HMM proteins were highly expressed (Terrat et al., 2012). Some of these proteins were also detected in the *C. consors* DV. To determine from which tissue they originate, the levels of expression of these HMM proteins in the venom duct were compared with those obtained from the salivary gland EST databases by the number of Blast hits. The major peaks of expression of HMM proteins showed an interesting pattern (Table 17). The expression levels of PDI, PPI, and AK in the venom duct and the salivary gland were similar. Hyaluronidase and conodipine were expressed more strongly in the venom duct, and echotoxin more in the salivary gland (Table 17). For comparison, LMM conotoxins and conkunitzins were expressed preferentially in the venom duct (Table 17). The analysis of transcriptomic data

therefore clearly showed that not all the HMM proteins in the DV are synthesized in the venom duct but that some are coming from the salivary glands as well (Biggs et al., 2008).

Table 17: Difference in expression levels of genes encoding HMM venom proteins between venom duct and salivary gland

expressed gene	<i>C. consors</i> CC8 (VD + SG)	<i>C. consors</i> CC6 (VD + SG)	<i>C. consors</i> VD6	<i>C. consors</i> SG6
PDI	20	20	15	19
PPI	22	17	9	12
AK	11	9	8	7
Hyaluronidase	40	300	260	1
Echotoxin	230	130	4	190
Conodipine	2	2	2	0
Conotoxin A superfamily (Cn4.4, 7/2)	299	568	569	2
Conotoxin A superfamily (Cn1.1, 3/5)	128	437	401	7
Conotoxin O1 superfamily	90	180	155	5
Conotoxin M superfamily	48	106	100	5
Conkunitzin	16	28	22	0

SG6: salivary gland transcriptome, VD6: venom duct transcriptome, CC6 and CC8: combined salivary gland and venom duct transcriptomes. These databases are deposited in the Conco database site (<http://conco.ebc.ee>). Significant difference in tissue-specific expression of some genes is emphasized in bold.

4. Conclusions

The first comprehensive analysis of the HMM proteins of the dissected venom (DV) and the injected venom (IV) of the fish-hunting cone snail *C. consors* was performed by an integrative approach, i.e., by matching proteomic data with nucleic acid information from the same species. Representatives of 22 different protein families have been identified in the *C. consors* DV and IV. Of these, orthologs of 10 *C. consors* venom HMM proteins have so far been described as being extracellular. Venom proteins with potential proteinase, proteinase inhibitory, peptidyl-prolyl bond cis-trans isomerization, disulfide bond isomerization, and arginine kinase activities may participate in venom maturation in the lumen of the venom duct, probably via the venom granules. Others having potential selective conotoxin-binding ability, such as PDI and AChR $\alpha 7$ subunit-like protein, may participate in selecting venom components to be injected. HMM proteins hyaluronidase and echotoxin, found in the IV, and possibly also sPLA2 and AK that have not been detected in the IV, may contribute to envenomation or digestion processes.

Maturation of toxin precursors is a multistep process that involves molecular chaperones and enzymes operating during their translocation across the ER membrane. Some post-translation processes in animal peptide toxins, such as proteolytic processing, C-terminal amidation, and disulfide bond formation, are found in a broad variety of secreted peptides (Buczek et al., 2005). The importance of protein post-translational processing in the context of the *C. consors* venom proteome is emphasized by the presence of transcripts encoding for proteins involved in the secretory pathway and post-translational processing, as reported for *C. bullatus* (Hu, Bandyopadhyay, Olivera, & Yandell, 2011). This and the previous analyses of transcriptomes and proteomes of venomous cone snails (Biass et al., 2009; Safavi-Hemami et al., 2011; Safavi-Hemami et al., 2010; Terrat et al., 2012; Hu et al., 2011) and other venomous animals have shown that their HMM venom proteins are mostly involved in

the correct folding, post-translational modification of conopeptides (e.g., in *C. bullatus* (Hu et al., 2011)), the cell defense system, and protection of venom toxins against oxidative stress.

It is particularly interesting that the new conoprotein families that have been found in the venom of *C. consors*, some highly represented, appear to be present exclusively in cone snails. Complete or even partial sequences of these structurally unique cone snail proteins will enable their function to be studied further. Also, differential expression sites of *C. consors* venom HMM proteins have been shown by analyzing venom duct and salivary gland EST libraries. Overall, these results have laid the basis for determining the precise subcellular localization of the identified venom protein and for understanding their function, especially at the venom duct and venom granule interfaces. This should eventually provide further insight into the molecular mechanisms involved in the complex biological venom system.

Acknowledgments

We acknowledge the financial support of the European Commission. This study has been performed as part of the CONCO cone snail genome project for health (<http://www.conco.eu>) within the 6th Framework Program (LIFESCI-HEALTH-6 Integrated Project LSHB-CT-2007, contract number 037592). Our sincere thanks go to Dr. Roger H. Pain for critical reading of the manuscript.

References

- Alsop, D. W. & Bettini, S. (1978). Arthropod venoms. In S.Bettini (Ed.), *Handbook of experimental pharmacology: New series* (pp. 977). Springer-Verlag: Berlin.
- Balbin, J. A. (1980). *Acetylcholinesterase of conus geographus venom: partial purification and characterization*. MS Thesis, Dept. of Biochemistry and Molecular Biology, University of the Philippines, Manila, Philippines.
- Bias, D., Dutertre, S., Gerbault, A., Menou, J. L., Offord, R., Favreau, P. et al. (2009). Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J.Proteomics.*, 72, 210-218.
- Biggs, J. S., Olivera, B. M., & Kantor, Y. I. (2008). Alpha-conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon*, 52, 101-105.
- Buczek, O., Bulaj, G., & Olivera, B. M. (2005). Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol.Life Sci.*, 62, 3067-3079.
- Bulaj, G., Buczek, O., Goodsell, I., Jimenez, E. C., Kranski, J., Nielsen, J. S. et al. (2003). Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. *Proc.Natl.Acad.Sci.U.S.A*, 100, 14562-14568.
- Castellanos-Serra, L., Vallin, A., Proenza, W., Le Caer, J. P., & Rossier, J. (2001). An optimized procedure for detection of proteins on carrier ampholyte isoelectric focusing and immobilized pH gradient gels with imidazole and zinc salts: its application to the identification of isoelectric focusing separated isoforms by in-gel proteolysis and mass spectrometry analysis. *Electrophoresis*, 22, 1677-1685.

- Chang, D. & Duda, T. F., Jr. (2012). Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol.Biol.Evol.*, 29, 2019-2029.
- Czerwicz, E., Begley, G. S., Bronstein, M., Stenflo, J., Taylor, K., Furie, B. C. et al. (2002). Expression and characterization of recombinant vitamin K-dependent gamma-glutamyl carboxylase from an invertebrate, *Conus textile*. *Eur.J.Biochem.*, 269, 6162-6172.
- Dutertre, S., Biass, D., Stocklin, R., & Favreau, P. (2010). Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon*, 55, 1453-1462.
- Favreau, P., Krimm, I., Le, G. F., Bobenrieth, M. J., Lamthanh, H., Bouet, F. et al. (1999). Biochemical characterization and nuclear magnetic resonance structure of novel alpha-conotoxins isolated from the venom of *Conus consors*. *Biochemistry*, 38, 6317-6326.
- Gertz, E. M., Yu, Y. K., Agarwala, R., Schaffer, A. A., & Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.*, 4, 41.
- Girish, K. S. & Kemparaju, K. (2007). The magic glue hyaluronan and its eraser hyaluronidase: a biological overview. *Life Sci.*, 80, 1921-1943.
- Gunji, K., Ishizaki, S., & Shiomi, K. (2010). Cloning of complementary and genomic DNAs encoding echotoxins, proteinaceous toxins from the salivary gland of marine gastropod *Monoplex echo*. *Protein J.*, 29, 487-492.

- Hansson, K., Thamlitz, A. M., Furie, B., Furie, B. C., & Stenflo, J. (2006). A single gamma-carboxyglutamic acid residue in a novel cysteine-rich secretory protein without propeptide. *Biochemistry*, 45, 12828-12839.
- Hopkins, C., Grilley, M., Miller, C., Shon, K. J., Cruz, L. J., Gray, W. R. et al. (1995). A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J.Biol.Chem.*, 270, 22361-22367.
- Hu, H., Bandyopadhyay, P. K., Olivera, B., & Yandell, M. (2011). Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC.Genomics*, 12, 60.
- Kawashima, Y., Nagai, H., Ishida, M., Nagashima, Y., & Shiomi, K. (2003). Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. *Toxicon*, 42, 491-497.
- Kemparaju K., Girish K.S., & Nagaraju S. (2010). Hyaluronidases, a Neglected Class of Glycosidases from Snake Venom: Behind a spreading factor. In S.P.Mackessy (Ed.), *Handbook of Venoms and Toxins of Reptiles* (pp. 237-258). CRC Press: Boca Raton.
- Khoudoli, G. A., Porter, I. M., Blow, J. J., & Swedlow, J. R. (2004). Optimisation of the two-dimensional gel electrophoresis protocol using the Taguchi approach. *Proteome Sci.*, 2, 6.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680-685.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.

- Mandel, R., Ryser, H. J., Ghani, F., Wu, M., & Peak, D. (1993). Inhibition of a reductive function of the plasma membrane by bacitracin and antibodies against protein disulfide-isomerase. *Proc.Natl.Acad.Sci.U.S.A*, 90, 4112-4116.
- Marsh, H. (1971). The caseinase activity of some vermivorous cone shell venoms. *Toxicon*, 9, 63-67.
- Marx, U. C., Daly, N. L., & Craik, D. J. (2006). NMR of conotoxins: structural features and an analysis of chemical shifts of post-translationally modified amino acids. *Magn Reson.Chem.*, 44, S41-S50.
- McIntosh, J. M., Ghomashchi, F., Gelb, M. H., Dooley, D. J., Stoehr, S. J., Giordani, A. B. et al. (1995). Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*. *J.Biol.Chem.*, 270, 3518-3526.
- Miljanich, G. P. (2004). Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr.Med.Chem.*, 11, 3029-3040.
- Milne, T. J., Abbenante, G., Tyndall, J. D., Halliday, J., & Lewis, R. J. (2003). Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J.Biol.Chem.*, 278, 31105-31110.
- Miranda, R. R. (1982). *Partial purification and characterization of Conus textile venom phosphodiesterase I*. MS Thesis, Dept. of Biochemistry and Molecular Biology, University of the Philippines, Manila, Philippines.
- Olivera, B. M. (2006). *Conus peptides: biodiversity-based discovery and exogenomics*. *J.Biol.Chem.*, 281, 31173-31177.
- Petersen, T. N., Brunak, S., von, H. G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, 8, 785-786.

- Qian, J., Guo, Z. Y., & Chi, C. W. (2008). Cloning and isolation of a conus cysteine-rich protein homologous to Tex31 but without proteolytic activity. *Acta Biochim.Biophys.Sin.(Shanghai)*, 40, 174-181.
- Safavi-Hemami, H., Bulaj, G., Olivera, B. M., Williamson, N. A., & Purcell, A. W. (2010). Identification of Conus peptidylprolyl cis-trans isomerases (PPIases) and assessment of their role in the oxidative folding of conotoxins. *J.Biol.Chem.*, 285, 12735-12746.
- Safavi-Hemami, H., Gorasia, D. G., Steiner, A. M., Williamson, N. A., Karas, J. A., Gajewiak, J. et al. (2012). Modulation of Conotoxin Structure and Function Is Achieved through a Multienzyme Complex in the Venom Glands of Cone Snails. *J.Biol.Chem.*, 287, 34288-34303.
- Safavi-Hemami, H., Siero, W. A., Gorasia, D. G., Young, N. D., Macmillan, D., Williamson, N. A. et al. (2011). Specialisation of the venom gland proteome in predatory cone snails reveals functional diversification of the conotoxin biosynthetic pathway. *J.Proteome.Res.*, 10, 3904-3919.
- Safavi-Hemami, H., Young, N. D., Williamson, N. A., & Purcell, A. W. (2010). Proteomic interrogation of venom delivery in marine cone snails: novel insights into the role of the venom bulb. *J.Proteome.Res.*, 9, 5610-5619.
- Schaloske, R. H. & Dennis, E. A. (2006). The phospholipase A2 superfamily and its group numbering system. *Biochim.Biophys.Acta*, 1761, 1246-1259.
- Shiomi, K., Kawashima, Y., Mizukami, M., & Nagashima, Y. (2002). Properties of proteinaceous toxins in the salivary gland of the marine gastropod (*Monoplex echo*). *Toxicon*, 40, 563-571.

- Stanley, T. B., Stafford, D. W., Olivera, B. M., & Bandyopadhyay, P. K. (1997). Identification of a vitamin K-dependent carboxylase in the venom duct of a *Conus* snail. *FEBS Lett.*, 407, 85-88.
- Stiles, B. G. (1993). Acetylcholine receptor binding characteristics of snake and cone snail venom postsynaptic neurotoxins: further studies with a non-radioactive assay. *Toxicon*, 31, 825-834.
- Tastet, C., Lescuyer, P., Diemer, H., Luche, S., Van, D. A., & Rabilloud, T. (2003). A versatile electrophoresis system for the analysis of high- and low-molecular-weight proteins. *Electrophoresis*, 24, 1787-1794.
- Tautz, D. & Domazet-Loso, T. (2011). The evolutionary origin of orphan genes. *Nat Rev.Genet.*, 12, 692-702.
- Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stocklin, R. et al. (2012). High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon*, 59, 34-46.
- Tu, A. T. & Hendon, R. R. (1983). Characterization of lizard venom hyaluronidase and evidence for its action as a spreading factor. *Comp Biochem.Physiol B*, 76, 377-383.
- Violette, A., Biass, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F. et al. (2012a). Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J.Proteomics.*, 75, 5215-5225.
- Violette, A., Leonardi, A., Piquemal, D., Terrat, Y., Biass, D., Dutertre, S. et al. (2012b). Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar.Drugs*, 10, 258-280.

Yamamoto, T., Arimoto, H., Kinumi, T., Oba, Y., & Uemura, D. (2007). Identification of proteins from venom of the paralytic spider wasp, *Cyphononyx dorsalis*. *Insect Biochem.Mol.Biol.*, 37, 278-286.

6.6 Concluding remarks

This study provides the first extensive analysis of high molecular mass components found in the dissected and injected venoms of a piscivorous cone snail. While some of these proteins were also detected in the injected venom as a contributing factor in the envenomation process (as explained in the injected venom study above), all the other proteins were only present in the dissected venom. This is not surprising as many proteins and enzymes are involved in the maturation process of the venom. Indeed, conopeptides undergo many post-translational modifications and enzymes play therefore an essential role in this post-translational processing. Many proteins with potential venom maturation activities have been identified, such as proteinases, proteinase inhibitors, disulfide bond isomerases, etc. Their activity in the maturation process is certainly achieved in the lumen of the venom duct, but their precise location all along the venom duct is still uncertain. The venom gland granules are thought to play an important role in this maturation process as their abundance is not linear along the venom duct as previously reported ¹⁸. This maturation process is further discussed in the following study and new hypotheses are formulated to try and understand why certain components are found in the dissected venom and not in the injectable venom.

Uncovering intense protein diversification in a cone snail venom gland using an integrative venomomics approach

Daniel Biass¹, Aude Violette¹, Nicolas Hulo¹, Frédérique Lisacek^{2,3}, Philippe Favreau¹ and Reto Stöcklin¹

¹ Atheris Laboratories, case postale 314, CH-1233 Bernex-Geneva, Switzerland

² Proteome Informatics Group, Swiss Institute of Bioinformatics, rue Michel Servet 1, CH-1211 Geneva 4, Switzerland

³ Section of Biology, University of Geneva, CH-1211 Geneva 4, Switzerland

Publication data

Published the 23.12.2014 in the Journal of Proteome Research, Vol. 14, Issue 2, p.628-638.

Pubmed ID: 25536169

DOI: 10.1021/pr500583u

Keywords

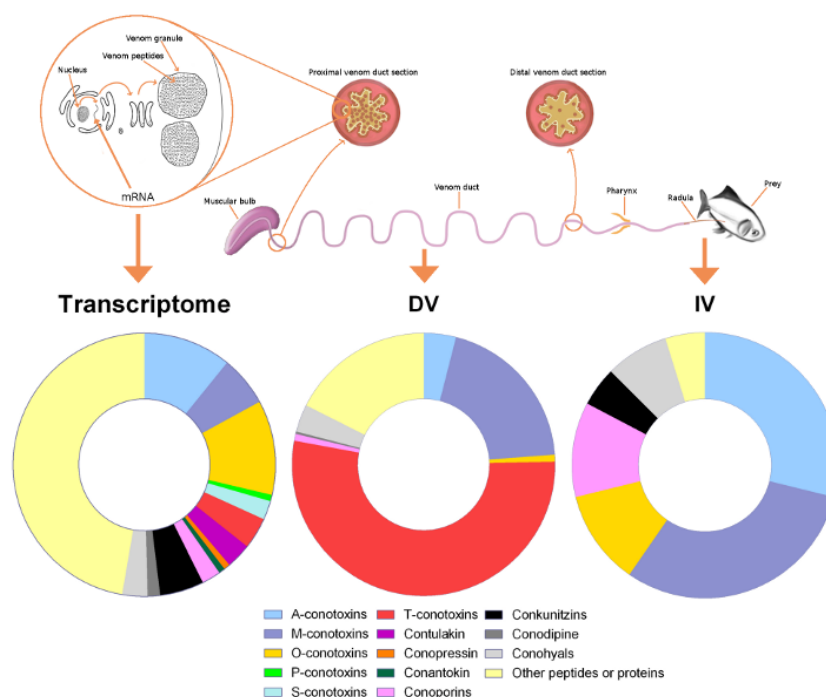
Cone snail, conopeptide, *Conus consors*, nanotips emitter, ESI-MS, mass spectrometry, peptide, peptidomics, transcriptomics, venomomics.

Personal contribution

All the work from sampling to data analysis and the manuscript writing.

Abstract

Marine cone snail venoms are highly complex mixtures of peptides and proteins. They have been studied in-depth over the past three decades, but the *modus operandi* of the venomous apparatus still remains unclear. Using the fish-hunting *Conus consors* as a model, we present an integrative venomomics approach, based on new proteomic results from the venom gland and data previously obtained from the transcriptome and the injectable venom. We describe here the complete peptide content of the dissected venom by the identification of numerous new peptides using nanospray tandem mass spectrometry in combination with transcriptomic data. Results reveal extensive mature peptide diversification mechanisms at work in the venom gland. In addition, by integrating data from three different venom stages - transcriptome, dissected, and injectable venoms – from a single species, we obtain a global overview of the venom processing that occurs from the venom gland tissue to the venom delivery step. In the light of the successive steps in this venom production system, we demonstrate that each venom compartment is highly specific in terms of peptide and protein content. Moreover, the integrated investigative approach discussed here could become an essential part of pharmaceutical development as it provides new potential drug candidates, and opens the door to numerous analogues generated by the very mechanisms used by nature to diversify its peptide and protein arsenal.



1. Introduction

Cone snails are predatory marine mollusks of the *Neogastropoda* clade (Bouchet et al., 2005), with more than 750 species described to date. Nonetheless, new species are discovered on a regular basis (Puillandre et al., 2014). Through millions of years of evolution, species of the monophyletic *Conoidea* superfamily (=Toxoglossa) have diverged from this clade and evolved a highly sophisticated venom apparatus for prey capture and/or defense purposes (Puillandre et al., 2008; Taylor, Kantor, & Sysoev, 1993). Furthermore, the different morphologies of their digestive system and venom apparatus have enabled the development of different feeding mechanisms that define subfamilies (Taylor et al., 1993). In one of these subfamilies, the *Conidae*, to which the cone snails belong, three different feeding habits have been characterized – vermivorous (feeding on worms), molluscivorous (feeding on other mollusks) and piscivorous (fish hunters) – according to the different morphologies of one of the most important features of the venom apparatus – the radula tooth. This harpoon-shaped weapon can be compared to a hypodermic needle, as it enables the rapid and efficient delivery of the venom into the prey (Salisbury, Martin, Kier, & Schulz, 2010).

Cone snail venoms are complex mixtures of highly active bio-compounds. Depending on the species and type of sampling, the number of venom components ranges from a hundred to thousands of molecules (Biass et al., 2009; Violette et al., 2012a; Safavi-Hemami et al., 2011), commonly known as conopeptides. Many of them referred to as conotoxins bear one or more disulfide bridges that confer a highly rigid structure. Typically, conopeptides are 10 to 35 amino acids long and traditionally classified into families according to their biological activity. Recent estimates of cone snail species and venom component numbers (Biass et al., 2009; Dutertre et al., 2013), and statistics of conopeptides described in UniProtKB or the Conoserver database (Kaas, Westermann, Halai, Wang, & Craik, 2008), suggest that

conopeptide diversity is still very much underestimated. Approximately 1,000 species each with a unique set of minimally hundreds of conopeptides, including intra-species and even intra-specimen variations (Dutertre, Biass, Stocklin, & Favreau, 2010), generates a natural library breaching the one million barrier, which is significantly larger than the 2,500 conopeptides described to date.

Serious cases of human envenomation by piscivorous cone snails led to the intensive study of this group and their dangerous venom (Flecker, 1936). Initiated in the late 60's, cone snail venom research led in the mid 80's to the discovery of the omega-conotoxin MVIIA (Olivera et al., 1985), a *Conus magus* venom conotoxin that selectively blocks N-type calcium channels and was approved by the FDA in 2004 for the treatment of severe chronic pain (Miljanich, 2004). Other molecules are currently in different stages of the drug development pipeline for the treatment of many neurological diseases or neuropathic disorders (Alonso, Khalil, Satkunanthan, & Livett, 2003; Lewis, Dutertre, Vetter, & Christie, 2012), as direct or indirect treatments for cancers (Dave & Lahiry, 2012; Luo et al., 2013; Kraus et al., 2011; Abdel-Rahman, Abdel-Nabi, El-Naggar, Abbas, & Strong, 2013). Unfortunately, conventional bioactivity-guided discovery approaches hitherto followed are time-consuming requiring usually several fractionation steps from the raw venom and substantial amounts of biological samples for full amino acid sequencing using biochemical techniques. This drastically limits the discovery of novel components and thus explain why far less than 1%, of conopeptide diversity has been structurally characterized to date (Puillandre, Koua, Favreau, Olivera, & Stocklin, 2012).

This situation boosted the emergence of venomomics that involves systems biology studies of venomous animals towards understanding the venomous function and applying this knowledge to drug discovery and to the development of pharmacological research tools (Menez, Stocklin, & Mebs, 2006).

Paradoxically, with the advent of so-called second- or next-generation sequencing platforms, recent venom gland transcriptomic studies (RNA-Seq) revealed “only” about a hundred gene isoforms coding for potential conopeptides in each species (Dutertre et al., 2013; Hu, Bandyopadhyay, Olivera, & Yandell, 2011; Lluisma, Milash, Moore, Olivera, & Bandyopadhyay, 2012; Terrat et al., 2012; Jin et al., 2013). At the transcript level, the signal sequence and propeptide region of mRNAs coding for conopeptides generally show a high level of conservation within a given conopeptide superfamily. In contrast, the potential conopeptide mature sequence of the mRNA is highly variable (except for the number or position of the cysteine residues that confer the tertiary structure to the peptide), providing the cone snail with a few isoforms of certain conopeptide superfamily and family. The evolution and variability of the mature conopeptide sequence in the transcriptome could hypothetically arise from a very quick diversification of the genes due to selective splicing, exon shuffling, irregular crossover or specific recombination at the genomic level (Pi et al., 2006). This may also result from high frequency gene duplication in conjunction with nucleotide shifts or substitutions in the mature peptide loci creating a diversifying and selective mechanism (Duda, Jr. & Palumbi, 1999; Conticello et al., 2001). A recent study proposed a deficiency or down-regulation of a mismatch repair system in the variable mature part of the peptide sequence in a given exon, and a site-specific recombination mechanism as additional though hypothetical diversification processes (Olivera et al., 1999). The discrepancy between the number of gene isoforms found at transcriptomic level and the number of mature conopeptides at the proteomic level suggests that the peptide maturation and post-translational modifications (PTMs) also play an unequivocal role in the final mature peptide processing, thereby producing the high level of venom diversity in different cone snail species (Safavi-Hemami et al., 2011; Dutertre et al., 2013; Safavi-Hemami et al., 2012).

Expectedly, venom sampling methods influence conotoxin identification: the peptide content of injectable venom (IV) (obtained by milking live animals and also called “milked venom” or “injected venom” in literature) and dissected venom (DV) widely varies. Moreover, striking

changes in venom composition may occur for instance due to intra-species and intra-specimen variations with time or with type of stimuli used for collection, and further complicates the study of venoms (Dutertre et al., 2010; Jakubowski, Kelley, Sweedler, Gilly, & Schulz, 2005; Dutertre et al., 2014). Venom complexity is further confirmed by the use of a broad array of analytical instruments providing complementary results (Biass et al., 2009; Dutertre et al., 2013).

A fully integrated approach that combines genomics, transcriptomics and peptidomics/proteomics therefore seems essential for understanding venom production, venom maturation and envenomation mechanisms. To this end, the extensive study of the *Conus consors* (also called *Pionoconus consors* in new classification (Tucker & Tenorio, 2009)) was proposed as part of the CONCO European consortium project (the cone snail genome project for health). Here we describe the peptide composition of the dissected venom of *Conus consors* using combined liquid chromatography and mass spectrometry analysis and the broad analysis of the venom duct transcriptome using specifically developed bioinformatics tools. We also present the first full comparative analysis of the conopeptide and protein content of dissected and injectable venoms, based on published datasets. We finally give an overview of the different steps of venom production and delivery occurring in a cone snail venom gland, highlighting extensive and unexpected diversification mechanisms.

2. Materials and methods

Acetonitrile (ACN, from Fisher Scientific Ltd., Loughborough, UK), trifluoroacetic acid (TFA, from Thermo scientific, Rockford, IL, USA) and formic acid (FA, from Acros Organics, Geel, Belgium) were of HPLC gradient grade or higher. Dithiothreitol (DTT), ammonium bicarbonate, iodoacetamide and all other reagents were of analytical grade or better and solutions freshly prepared prior to use. Deionised water was purified using a Milli-Q system (Millipore Corp., Billerica, MA, USA). Solvents were filtered and sonicated before use.

2.1 Venom preparation

All 25 specimens of *Conus consors* used in this study were collected from one colony in the Chesterfield Islands (New Caledonia) during the CONFIELD-I scientific expedition in June 2007. The pool of crude venom, referred to here as DV (dissected venom), was obtained after dissection of 19 *Conus consors* specimens using a method described previously (Favreau et al., 1999). The IV dataset used for the comparative analyses with DV and transcriptomic data was published earlier in Violette et al. (2012). For the DV peptidomic analyses, the lyophilized dissected venom pool was reconstituted at 1 mg/mL (protein content) in acidified water (0.1 % TFA) and desalted using solid-phase extraction onto a Sep-Pak Vac 35cc cartridge equilibrated in acidified water according to the manufacturer's instructions (Waters, Milford, MA, USA). Elution was performed with 70 % ACN in acidified water and the eluate freeze-dried under vacuum in a SpeedVac concentrator (ThermoSavant, Holbrook, NY, USA), then stored at -80 °C.

2.2 Venom duct transcriptome

The raw data used for the MS/MS matching analyses and transcriptome analysis was generated by Terrat *et al.* (2012). The materials and methods used by Terrat *et al.* are briefly reported here as follows. RNA was extracted from the venom ducts, previously stored in RNA-later (Qiagen), of three *Conus consors* specimens. To construct a cDNA library, 5 µg of total RNA was used. RNA quality was assessed in a Bioanalyzer 2100 (Agilent-Bonsai Technologies) and 5 µg of full-length double-stranded cDNA was then processed by the standard Genome Sequencer library-preparation method using the 454 DNA Library Preparation Kit (Titanium chemistry) to generate single-stranded DNA ready for emulsion PCR (emPCR™). The cDNA library was then nebulized according to the fragmentation process used in the standard Genome Sequencer shotgun library preparation procedure. The cDNA library was sequenced according to GS FLX technology (454/Roche). The short-

reads from the sequencing were assembled by MIRA version 2.9.25 using enhanced 454 parameters.

2.3 RP-HPLC

Reversed-phase high-performance liquid chromatography (RP-HPLC) was carried out using a Waters Alliance 2795 system equipped with a Waters 996 Photodiode Array Detector under control of the Waters Millennium³² 4.0 software (Waters, Milford, MA, USA). All fractionations were performed using a 218TP510 Protein and Peptide C₁₈ RP column (10 mm internal diameter / 250 mm length, from Vydac, Hesperia, CA, USA) with a gradient combining solvent A (0.1% TFA in water) and solvent B (90 % ACN / 0.1 % TFA in water). A flow rate of 2.0 mL/min was used with a gradient of 1 % B per minute, starting from 100% of solvent A. UV detection of the fractions was carried out at 214 nm and fractions were collected manually. The dissected venom pool was subjected to 5 RP-HPLC runs of 90 min each and corresponding fractions were pooled to obtain one set of fractions. A portion of each fraction was separated in aliquots for MS/MS analyses. All final fractions and aliquots were freeze-dried and stored at -80 °C.

2.4 Reduction of disulfide bonds and alkylation

Aliquots of lyophilized venom fractions were suspended in a 70 % milli-Q water / 30 % ACN solution at an approximate concentration of 1 mg/mL (value based on the UV-HPLC peak integration of each individual fraction). Ammonium bicarbonate and DTT were added to reach a final concentration of 2 % w/v and 4.5 mM respectively. The mixture was heated up to 60 °C for 2 h then cooled to ambient temperature and 10 µL of iodoacetamide (100 mM in water) was added for alkylation of free sulfhydryl groups and left to incubate at room temperature for 15 min, protected from light. Finally, 10 µL of a cysteine solution (200 mM in

water) was added and left to incubate at room temperature for 15 min to consume the iodoacetamide excess. Fractions were then freeze-dried and stored at -80 °C. Prior to mass spectrometry (MS) and tandem mass spectrometry (MS/MS) analysis, each fraction was acidified to pH 3.0 with FA 20 % and was desalted and concentrated using C18 phase ZipTip pipette tips (Millipore Corp., Billerica, MA, USA) using manufacturer's protocol except for elution. Fractions were directly eluted from the ZipTip pipette tips with 10 µL of solvent (H₂O/ACN/HCOOH 49.8:50:0.2, v/v/v) (with exception of fractions that eluted in RP-HPLC chromatogram after 50 % solution B as these were eluted from the ZipTip pipette tips with a solvent H₂O/ACN/HCOOH 29.8:70:0.2, v/v/v).

2.5 Nanotip emitter ESI-MS and MS/MS of venom fractions

Nano-ESI-MS and MS/MS analyses were performed on a Q-TOF Premier mass spectrometer (Waters-Micromass, Manchester, UK) equipped with a nano ESI source and operated under control of the MassLynx 4.1 software (Waters-Micromass). A few microliters of each fraction were deposited in a 0.58 mm internal diameter nanotip emitter with conductive coating (EconoTips Emitters, New Objective, Inc., Woburn, MA, USA). Acquisitions were carried out with a scan from 100 to 1800 *m/z* in 1 sec. in the positive ionization mode at an estimated flow rate of 50 nL/min. After MS analysis, the charge of each *m/z* signal detected was manually assessed and noted. Then MS/MS was manually carried out on each signal detected in a single fraction. The MassLynx software was used for data processing and to assist data analysis. External calibration of the instrument was performed with glu-fibrinopeptide-B with a maximum deviation tolerance of 20 ppm.

2.6 MS and MS/MS data analysis

All MS data were manually interpreted using the Biolynx module of MassLynx 4.1 software (Waters-Micromass). The MS/MS spectrum of each mass was deconvoluted into a singly-

charged m/z axis using the Maxent3 module (Waters-Micromass) according to the previously noted parent ion mass charge. Files were then submitted to Phenyx software version 2.6.2 (Genebio, Geneva, Switzerland – the scoring algorithm of the Phenyx software is detailed in (Colinge, Masselot, Giron, Dessingy, & Magnin, 2003)) and matched against the transcriptome database (all possible reading frames, see examples in Supplementary material 1). Parent mass tolerance was set to 0.2 Da and maximum p-value to 1.10^{-7} . Hydroxyproline, pyroglutamic acid, methionine oxidation, tryptophan bromination, glutamic acid carboxylation, tyrosine sulfation and C-terminus amidation were included as possible modifications. Carboxyamidomethyl cysteine was included as fixed modification. Matches were considered valid when z-score was greater than 6 (5.5 for <10 amino acids) and delta m/z less than 0.1. Matching results from MS/MS with the corresponding transcript were then manually assessed and checked using Biolynx and the PepSeq module (Waters-Micromass).

2.7 Bioinformatics

DV peptide sequences of interest were grouped by families and aligned with the MAFFT software (Katoh & Standley, 2013). An HMM-profile (Hidden Markov Model) was then built for each conotoxin family with the HMMer package (Eddy, 2011) and used to scan cone snail sequenced transcriptomes (unpublished data). Newly identified sequences were then automatically aligned with original sequences using ClustalW then manually re-aligned using the Jalview 2.8 software (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). The phylogenetic trees were generated with the SeaView 4 software (Gouy, Guindon, & Gascuel, 2010) using the PhyML program (Guindon et al., 2010) with the following options: JTT model of amino acids substitution (Jones, Taylor, & Thornton, 1992), BioNJ as initial tree, optimized tree topology and 1000 bootstraps.

3. Results and discussion

The aim of this study was to compare data of the *Conus consors* venom gland transcriptome (Terrat et al., 2012), the dissected venom content (this study, (Biass et al., 2009) and (Leonardi et al., 2012)) and the injectable venom content (Violette et al., 2012a) using an integrated approach. The dissected venom peptidome was also characterized using ESI-MS/MS techniques coupled to the transcriptome. This provides a global overview of the natural steps that occur in a venom apparatus of a piscivorous cone snail - from venom preparation to final delivery into the prey.

3.1 RP-HPLC

The first step of venom analysis involved an offline RP-HPLC of the pooled dissected venoms in order to fractionate venom components. Each peak was manually collected according to the 214 nm UV absorbance signal (see chromatogram in Figure 40), resulting in 108 fractions. Each run provided good peak resolution and reproducibility, enabling fraction pooling between runs. A portion of the fractions identical to those used for the MS comparative study (Biass et al., 2009) was then processed for reduction and subsequent MS and MS/MS analyses. This pre-analytical procedure enabled direct mass comparison between the native and reduced/alkylated components, allowing detection of any protein degradation and led to further MS and MS/MS analyses on individual fractions, when needed.

The dissected venom chromatogram displayed a far more complex pattern than for the injectable venom (see chromatograms on Figure 17 of the comparative MS study (Biass et al., 2009)). Interestingly, the dissected venom had two supplementary zones in the hydrophobic region of the chromatogram.

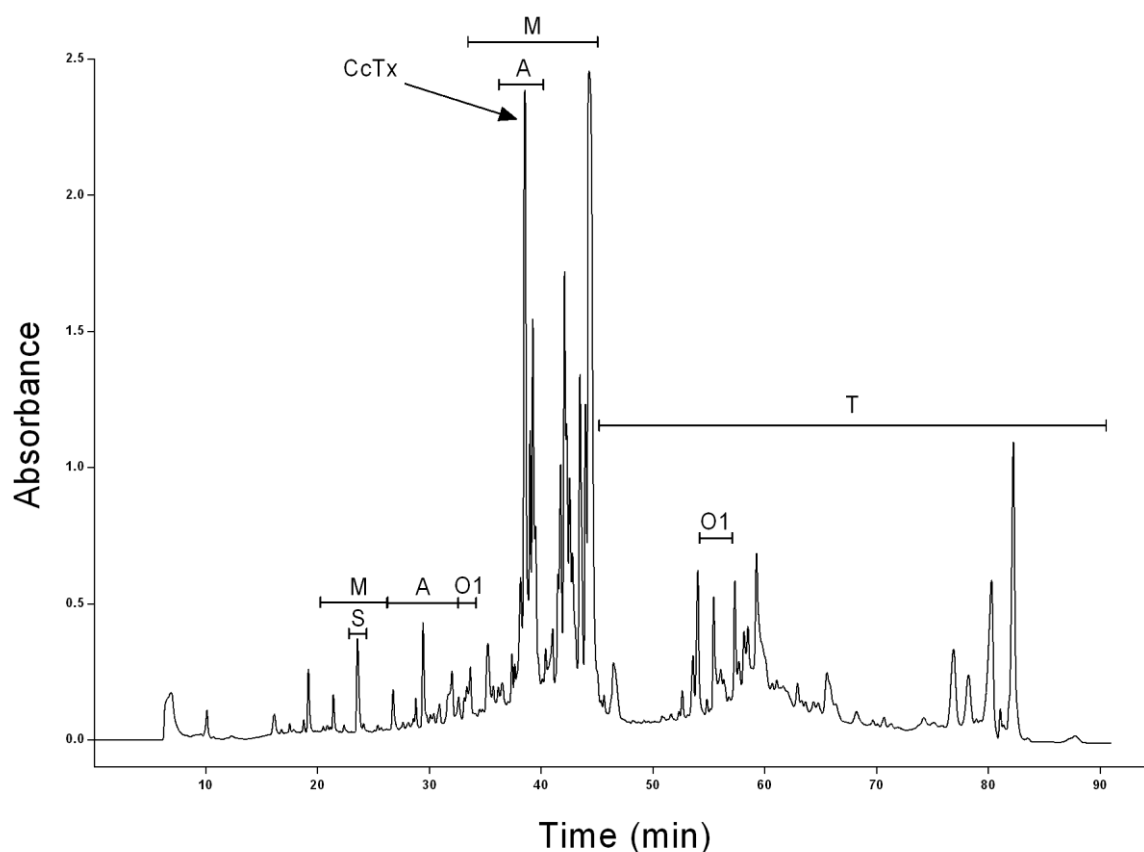


Figure 40: UV chromatogram at 214 nm obtained by RP-HPLC fractionation of the DV of *Conus consors*.

Peptide sequences of A-, M-, O-, S- and T-superfamilies are spread in different zones of the chromatogram.

3.2 MS analysis of dissected venom peptidome and comparison with injectable venom

Of the 108 native dissected venom fractions collected manually, 1078 masses were found by standard source ESI-MS (MS comparison study (Biass et al., 2009)); of the 80 native injectable venom fractions collected automatically (injectable venom study (Violette et al., 2012a)), 419 masses were found by LC-ESI-MS. This strong discrepancy in the number of components between IV and DV can be explained by the contribution of a specific activation mechanism. This hypothesis is supported by correlating IV and DV masses according to the

mass range distribution. The percentage of peptides in a given mass range over the total number of peptides in the IV was compared to that of the DV (Figure 41A). This revealed an increase of small components (500 to 1500 Da range) in the DV compared to the IV. This difference could stem from secondary products of the maturation process in the venom duct, such as small variable parts of propeptide sequences after mature peptide release in the venom granules or in the venom duct lumen. In contrast, the DV contained fewer components in the 3500-5000 Da range compared with the IV. This could be explained by the fact that some of the components present in the IV were not produced in the venom duct gland, but elsewhere in another gland, like the salivary gland as suggested by Biggs et al. (Biggs, Olivera, & Kantor, 2008). Additionally, the ratio of the number of IV and DV components for each mass range (Figure 41B), clearly revealed more components in the DV in ranges above 5000 Da, suggesting that some proteolytic and maturation processes occurred in the DV and resulted in a relative enrichment of peptide between 3500 and 5000 Da in the IV. This ratio comparison also demonstrates that peptides found in the DV in the lower mass ranges, between 500 to 3500 Da, have been naturally eliminated in the IV by potential selection mechanisms (Figure 41B). Lower and higher range masses that are substantially more present in the DV could also suggest potential signal saturation and ion suppression effects of the high and low masses over the medium masses during the analysis of each individual fraction, and in turn explain the lower signal level in the middle range. The offline analysis method of each individual fraction for the DV using nanospray tips in comparison to the online LC analysis used for the IV study may also account for this difference. The nanospray tips had the advantage of concentrating the samples but were more prone to ion suppression effects and higher signal-to-noise ratio due to uncontrollable flow rate and lack of liquid chromatographic dimension (Annesley, 2003). These results indicated a sharp change in venom composition between the DV and IV, implying either venom enrichment of particular peptides, protein processing or even an intermediate source

of products. We therefore concentrated on characterizing the protein content of each venom compartment to investigate the processes at work in the venom gland.

3.3 Peptide identification by MS/MS matching against transcriptome.

Each reduced and alkylated fraction was analyzed by offline nanotip ESI-MS and MS/MS. From the MS spectrum, each molecular species was manually assigned a mass charge and fragmented by gas collision induced dissociation. After deconvolution into a singly charged MS/MS spectrum, data were then processed with the Phenyx software to find corresponding transcriptomic sequences within a certain mass error. Common post-translational modifications (PTMs) observed in cone snail venoms were also taken into account in the parameters of the software during processing. Indeed, PTMs are present at high frequency and variability in this subfamily of gastropods (Buczek, Bulaj, & Olivera, 2005). In the peptidomic content of the DV, a total of 299 peptide sequences were clearly characterized using this integrated approach with transcriptomic data (see tables in Supplementary material 2). Among these identifications, 22 sequences could be assigned to A-superfamily transcripts, 82 to M-superfamily, 3 to O1-superfamily, 1 to S-superfamily and 191 to T-superfamily transcripts. The distribution of the retention time in the chromatogram (see Figure 40) showed distinct superfamily clusters.

In the A-superfamily (see Table 1 in Supplementary material 2), 1 sequence had the cysteine framework I and had been described previously (Violette et al., 2012a). All the α -conotoxins described previously (except for [Hyp4]-CnID, CnIJ and [Hyp7]-CnIK) were found using MS, but their sequences could not be deduced by transcriptomic matching due to low-quality MS/MS spectrum. Six propeptide region sequences belonging to α -conotoxin precursors were found, supporting the presence of these conotoxins in the DV. All of these conotoxins are known to be competitive antagonists of nicotinic acetylcholine receptors (nAChRs) and have been widely used as tools to discriminate between the different neuronal or muscular

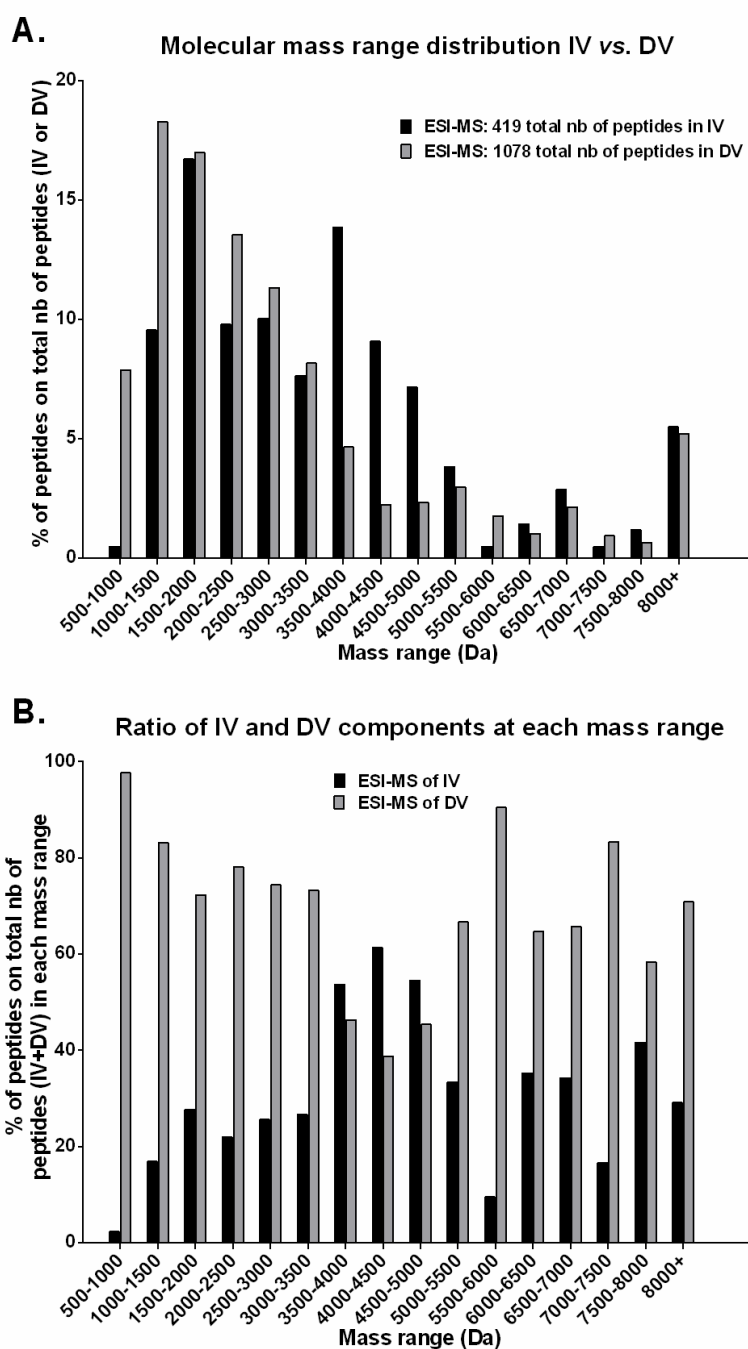


Figure 41: Molecular mass range distributions of native components found in the IV vs. DV.

Comparison of the percentage of peptides in a given mass range over the total number of peptides (A). Comparison of the percentage of peptides in a given mass range over the number of IV and DV peptides in the mass range (B).

types of nAChRs (Muttenthaler, Akondi, & Alewood, 2011). Fifteen other components belonging to the A-superfamily and presenting the $[-CC(X_7)C(X_2)CXC(X_3)C-]$ cysteine pattern were found. Of these, 8 sequences presenting different O-glycosylation and other PTM variations were related to the previously described CcTx compound (Le et al., 1999). This conotoxin is known to target neuronal voltage-gated sodium channels and has an excitatory effect on motor nerve terminals. The tertiary structure of this O-glycosylated component, which is one of the most abundant peptides present in both the IV and DV of *C. consors*, was only recently elucidated (Hocking et al., 2013), highlighting the difficulty in characterizing, and elucidating the structures of glycosylated peptides. The other 7 newly identified compounds with the cysteine IV framework also belonged, by homology, to the κ A-family, but were more closely related to the previously described κ A-MIVA than the CcTx. Mass spectral data of these components revealed glycosylation patterns similar to CcTx, but with a different combination of glycans. On the basis of the different sugar signals and by homology with κ A-MIVA, the deduced sugar composition revealed a duplication of a “1 N-acetyl-hexosamine - 2 hexoses” structure. The glycosylated sequences were identified and matched to the transcriptome using an in-house bioinformatic tool described previously (Violette et al., 2012b). By homology to the κ A-MIVA, these two O-glycosylations were tentatively placed on two threonines in positions 7 and 9 on all sequences (see Table 1 in Supplementary material 2). These κ A-conotoxins are known to contribute to excitotoxic shock during prey envenomation by possibly inhibiting potassium channels (Terlau et al., 1996; Teichert, Jacobsen, Terlau, Yoshikami, & Olivera, 2007). These peptides are very similar to previously identified *C. consors* sequences in genomic DNA and cDNA described by Puillandre et al. (Puillandre, Watkins, & Olivera, 2010). In addition, among these 15 A-superfamily components, 2 propeptide region sequences were also identified.

Concerning the M-superfamily (see Table 2 in Supplementary material 2), 7 sequences were identified with the cysteine framework III, most of which were previously characterized in the IV (Violette et al., 2012a). Within this cysteine pattern, 3 sequences contained only one

residue before the last cysteine pair (unique feature in a piscivorous cone snail). To date, no molecular target is known for these components. A number of M-superfamily conopeptides had no disulfide bridges (see Table 3 in Supplementary material 2); surprisingly, some of these, especially [Hyp6]-Conomarphin-Cn1 and [Hyp9]-Conomarphin-Cn24, were among the most abundant peptides present in the DV (see highest peak in the chromatogram on Figure 40). The signal and propeptide region sequences of these peptides had very high similarity to those of the conomarphin family, like the conomarphin-Mr1 (UniProt: B2KPN7), and the conopeptide Ac3.1 (UniProt: P0CH24). To confirm the assignment of these peptides in the conomarphin family, profiles were built from the two complete transcripts of Conomarphin-Cn1 and Conomarphin-Cn24 (see Table 3 in Supplementary material 2), and used to retrieve possible homologues from other cone snail transcriptomes. Similar sequences were found, some with higher phylogenetic commonalities like those of *Conus bullatus* and other more distant like those of *Conus eburneus* and *Conus sponsalis*. A cladogram tree was built by alignment of the mature part of these sequences, as well as with all sequences found in the Conoserver database (Kaas et al., 2008) in the M-superfamily, including the existing conomarphins. The tree clearly showed that the conomarphin family was, as expected, distinct from the cysteine-rich members of the M-superfamily, and that the new sequences, Conomarphin-Cn1 and Conomarphin-Cn24, could well be part of the conomarphin family (see Supplementary material 3). An alignment and a tree with only the new and previous members of the conomarphin family are shown in Figure 42. Although they do not have a cysteine framework, which makes alignments more difficult without an established backbone, the tree shows three sub-groups in this family. Conomarphins have no biological target assigned and their function in the DV remains unknown (Han et al., 2008).

In the O1-superfamily (see Table 4 in Supplementary material 2), 3 sequences with the VI/VII cysteine framework were identified. These were variants of previously identified compounds. Corresponding masses of many other components of this group (CnVIA to D, CnVIIA to D

and CnVIIJ) were found, but MS/MS spectra did not allow proper transcriptome matching, although the transcripts were present.

Concerning the S-superfamily, a single short propeptide sequence was found (see Table 5 in Supplementary material 2). Surprisingly, this peptide was identified with a pyroglutamic acid modification, showing that post-translational modifications are not restricted to mature peptides. However, neither mature sequences nor corresponding masses of this superfamily could be formally identified in the DV.

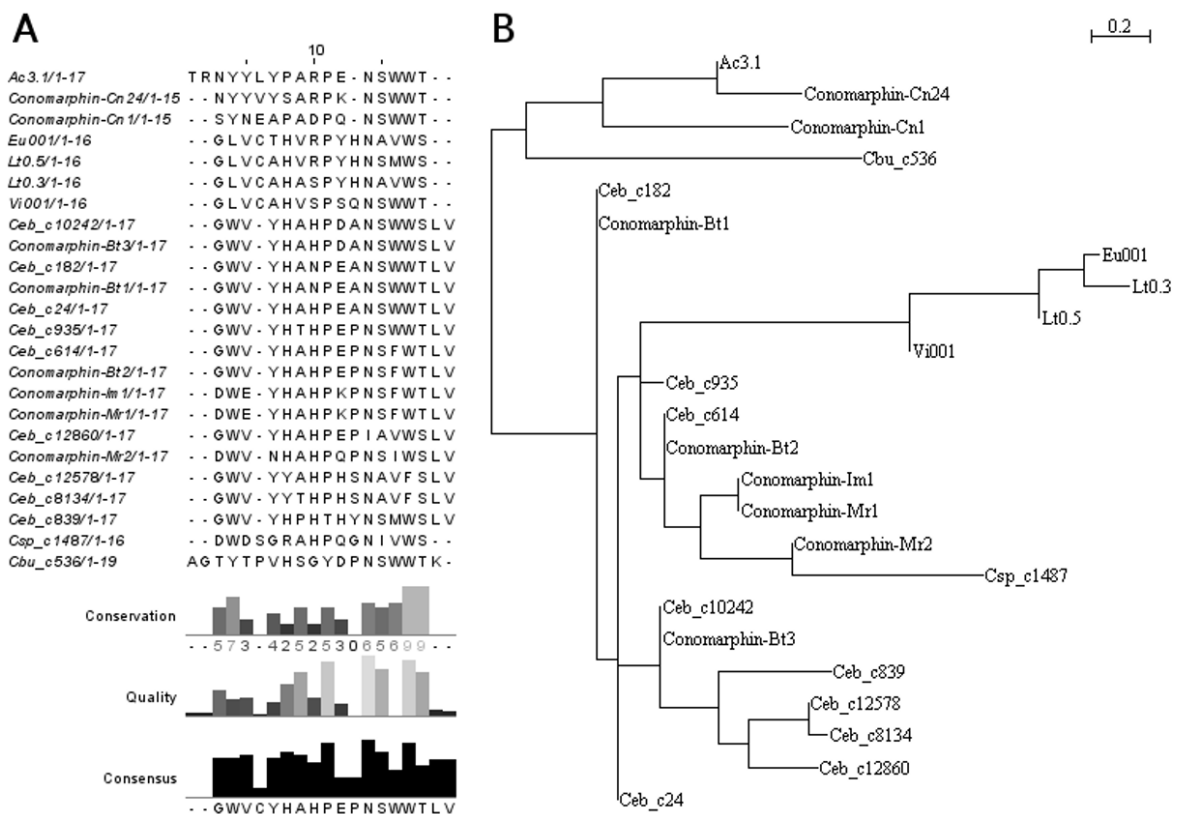


Figure 42: Alignment (A) and tree (B) of all reported and new conomarphins.

Finally, the major group of components found in the hydrophobic portion of the DV (see chromatogram on Figure 40) – that does not appear in the IV (see chromatogram on Figure 43 of the comparative MS study (Biass et al., 2009)) – mostly belonged to the T-superfamily.

All sequences found in this superfamily contained no disulfide bridges, and could be separated into two groups (see Table 6 and 7 in Supplementary material 2). Profiles were also tested against our in-house transcriptomic databases, but no significant sequences matches were found. A unique homologous sequence (P01107 in Conoserver database) was found in *Conus striatus*, a cone snail phylogenetically close (same clade) to *Conus consors* (Bandyopadhyay et al., 2008). The fact that these sequences were present only in species of the same clade can be explained by the rapid evolution of the genes in this peptide superfamily. Although these peptides are present in relatively high amounts in the DV, their biological activity is not currently known. A role as peptide modulators of an enzyme or receptor could be suggested.

3.4 Conopeptide precursors under multi-processing: mechanisms for diversity

A major result of this work is the exceptionally high number of different and overlapping conopeptides expressed from single mRNA transcripts (Table 18). This is illustrated by the M- and T-superfamilies, with for example, up to 44 peptides deriving from a single T-superfamily transcript. The length of the variable part of the C- and N-termini notably differed depending on the presence/absence of a cysteine framework in the conopeptide. Length variability of C- or/and N-termini was much higher in linear conopeptides than in disulfide-rich conopeptides. Disulfide bridges confer a strong tertiary structure that may limit the processing of the mature or propeptide sequences into different truncated forms. This difference between disulfide-rich and disulfide-poor peptides could mean that the maturation process occurs after the refolding process in venom production. Although artifactual modifications due to extraction cannot be totally excluded, this result is in line with previous observations made on the injectable venom of fish-hunting (Violette et al., 2012a) and mollusk-hunting (Dutertre et al., 2013) species. This would indicate that this variable peptide processing is not restricted to the injectable venom and would be widespread in the *Conus*

genus. As with the well-known and various post-translational modifications of conopeptides, this process could be part of a natural mechanism to pharmacologically enrich the venom cocktail for optimal survival (Dutertre et al., 2014; Violette et al., 2012a; Dutertre et al., 2013; Kaponi et al., 2013).

Table 18: Example of C- and N-terminus hyper-variability with DV conopeptides of the T-superfamily.

In the transcript sequence, the signal sequence is in blue, the propeptide region in green, the mature peptide in black and the stop codon is represented by a red “X”.

Identity code	Sequence
98_18369_F6	MLCLPVFIILLLLASSAAPNPLETRIQS DLIRADLEDADTQPDER FFSSI AKLVSKVAPAVADAIIPVITDLVNGA X
Cn48	ADLEDADTQPDERFFSSI AKLV
Cn49	FFSSI AKLVSKVAPAVADAIIPVITDLVNGA
Cn50	FFSSI AKLVSKVAPAVADAIIPVITD
Cn51	FFSSI AKLVSKVAPAVADAIIPVIT
Cn52	FFSSI AKLVSKVAPAVAD
Cn53	FFSSI AKLVSKVAPAVA
Cn54	FFSSI AKLV
Cn55	FSSI AKLV
Cn56	SSI AKLV
Cn57	SI AKLVSKVAPAVADAIIPVITDLVNGA
Cn58	SI AKLVSKVAPAVADAIIPVITD
Cn59	SI AKLV
Cn60	IA KLVS KVAPAVADAIIPVITDLVNGA
Cn61	IA KLVS KVAPAVADAIIPVITDLVN
Cn62	IA KLVS KVAPAVADA
Cn63	IA KLVS KVAPAVAD
Cn64	LV SKVAPAVADAIIPVITDLVNGA
Cn65	LV SKVAPAVADAIIPVITDLVNG
Cn66	LV SKVAPAVADAIIPVITDLVN
Cn67	LV SKVAPAVADAIIPVITDLVN
Cn68	LV SKVAPAVADAIIPVITDL
Cn69	LV SKVAPAVADAIIPVITD
Cn70	LV SKVAPAVADAIIPVIT
Cn71	LV SKVAPAVAD
Cn72	LV SKVAPA

Cn73	VSKVAPAVADAIIPVITDLVN
Cn74	KVAPAVADAIIPVITDLVNGA
Cn75	KVAPAVADAIIPVITDLVN
Cn76	KVAPAVADAIIPVITDL
Cn77	VAPAVADAIIPVITDLVNGA
Cn78	VAPAVADAIIPVITDLVN
Cn79	VAPAVADAIIPVITDL
Cn80	VAPAVADAIIPVITD
Cn81	VAPAVADAIIPVIT
Cn82	AVADAIIPVITDLVNGA
Cn83	AVADAIIPVITDLVN
Cn84	DAIIPVITDLVNGA
Cn85	DAIIPVITDLVN
Cn86	DAIIPVITDL
Cn87	AIPVITDLVNGA
Cn88	AIPVITDLVN
Cn89	IIPVITDLVNGA
Cn90	PVITDLVNGA
Cn91	PVITDLVN

3.5 From mRNA to prey: unexpected protein specificity along the venom production steps

Our data from the proteome of the venom duct were compared with the molecular data recently obtained from the transcriptome (Terrat et al., 2012) and the proteomic of *Conus consors* IV (Violette et al., 2012a). This provided an unprecedented dataset delineating the evolution of the venom composition from expression and synthesis to delivery (Figure 43C). Previous results showed that approximately 50% of the transcripts matched conopeptides or conotoxins (Terrat et al., 2012), and the rest was either common cellular genes, transposable elements, xenobiotics or unknown function genes (called “other peptides or proteins” in the Figure 43C Transcriptome). The proportion of conopeptide isoforms reflected the statistics established by the Conoserver for all *Conus* species (Kaas, Westermann, & Craik, 2010), in which precursors of the O-superfamily formed the major group, followed by the A-

superfamily, and the M- and T-superfamilies. In the DV, this hierarchy was upset, since clearly the peptides of the T-superfamily represented a predominant part of the pie chart (Figure 43C DV). This higher level of T-superfamily peptides could probably be explained by the fact that most of the peptides found in this group were linear and that, as explained above, the hyper-variability phenomenon was more pronounced with linear peptides. Surprisingly, this group is not present in the IV (Figure 43C IV). This could be caused by variations in the quantity of venom granules along the venom duct; indeed, the proximal part of the venom duct contained much more venom granules than the distal part (Figure 43B). This has also been observed in other cone snail species where the pharynx appears to play an important role in the final composition of the venom cocktail (Marshall et al., 2002). Another hypothesis is the specific use of T-superfamily peptides in the venom in reaction of a predator attack. Indeed, recent findings demonstrated that injected venoms may differ according to predation- or defense-stimuli (Dutertre et al., 2014). Also, these peptides could play a role in the production and maturation of conotoxins, for example by modulating the activity of the different processing enzymes. Indeed, enzymes, such as peptidyl-prolyl-isomerase (PPI) or protein disulfide isomerase (PDI), play an important role in maturation and processing of the venom components (Leonardi et al., 2012). A recent study demonstrated that the proteolytic activity of venom varies along the venom duct (Moller, Vanderweit, Bubis, & Mari, 2013). Proteolytic activity gradually increases from the muscular bulb towards the pharynx (Figure 43B), and finally disappears completely close to the pharynx, which clearly suggests an enzymatic activity modulation. This is in line with previous investigations performed by Jimenez et al. (Jimenez, Olivera, & Cruz, 1983) where proteolytic activities were located in venom granules present in the venom duct.

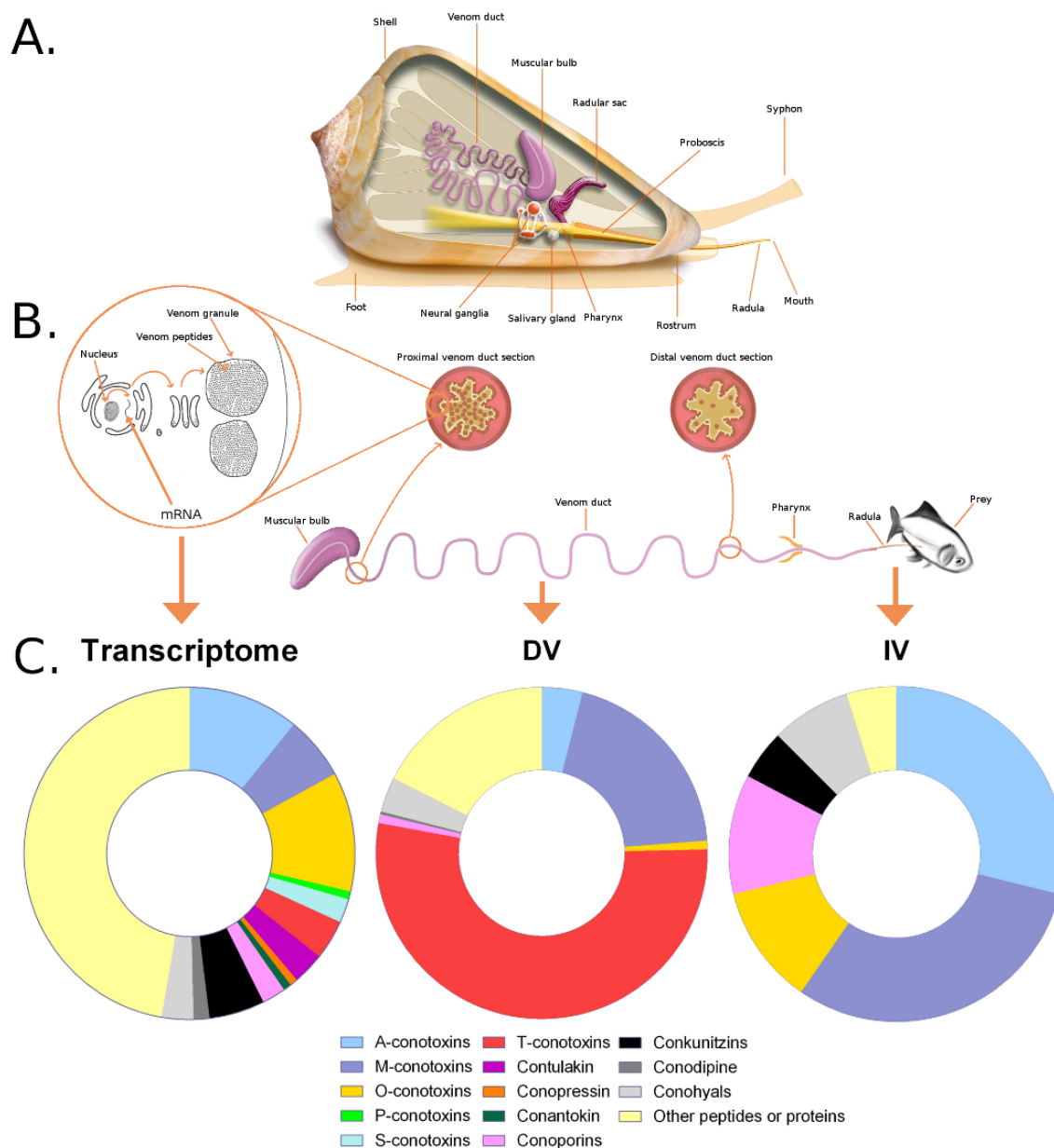


Figure 43: Overview of the cone snail general anatomy (A) and the venom apparatus from the elaboration of the venom and venom granules in the venom gland tissues to delivery to prey (B). Family distribution of conopeptides, conoproteins and other components in the transcriptome, the DV and the IV (C).

4. Conclusion

A total of 299 peptide sequences of the 1078 masses detected by mass spectrometry were fully elucidated by transcriptomic-assisted MS/MS data matching and could be assigned to existing A-, M-, O-, S- and T-superfamilies in the venom duct of *Conus consors*. New sequences belonging to known families, as well as completely new families in the T-superfamily, were found. The sequences in these new families bear no disulfide bridges and represent an important portion of the venom composition, possibly suggesting that these linear peptides may have been hitherto overlooked. This work on the *Conus consors* dissected venom also highlighted a high variability on both the C- and N-termini of conopeptide sequences. High variability was also observed in the IV and may endow the cone snails with a naturally enriched venom cocktail that offers wider envenoming properties. The fully integrated approach of this study not only led to the discovery of new families of components in the dissected venom, but also – and for the first time – provided an overview of venom processing by comparing results at three different levels in a single species: in the transcriptome of the venom gland, in the peptidome and proteome of the dissected venom and finally in the peptidome and proteome of the injectable venom. Unexpectedly, our results showed no clear correlation between mRNA expression, venom produced in the venom duct, and venom injected into the prey, highlighting the complex venomic machinery of a cone snail, and implying numerous mechanisms for protein expression modulation and transport as recently reported. Just as importantly perhaps, this integrated approach has furthered our understanding of the venom processing mechanism in cone snails that may turn out to be instrumental for further discoveries involving other species. With the advent of new and faster techniques in MS for peptidomics and next-generation sequencing for transcriptomics, combined with new high throughput biological assays, integrated approaches such as the present one could prove very useful for the discovery of new potential therapeutics.

Acknowledgements

We are most grateful to the European Commission for funding, as this study has been performed as part of the CONCO cone snail genome project for health (<http://www.toxinomics.org/conco/>) within the 6th Framework Program (LIFESCIHEALTH-6 Integrated Project LSHB-CT-2007, contract number 037592). This project is part of the “Venomics” initiative dedicated to the understanding of the function and evolution of venomous systems in various phyla (Menez et al., 2006). We would like to express our deepest gratitude to the Governments of New Caledonia and French Polynesia, the French Navy, the IRD-Nouméa (Fabrice Colin, Napoléon Colombani, Jean-Louis Menou and Claude Payri). The Toxinomics Foundation office in Nouméa (Alain Gerbault and Jacques Pusset) and Robin Offord (Mintaka Foundation for Medical Research in Geneva, Switzerland) for their constant support. We would also like to thank Cécile Cros and Frederic Perret from our laboratories for their help and Xavier Sprüngli from the Toxinomics Foundation (<http://www.toxinomics.org>) for cone snail scheme.

References

- Abdel-Rahman, M. A., Abdel-Nabi, I. M., El-Naggar, M. S., Abbas, O. A., & Strong, P. N. (2013). Conus vexillum venom induces oxidative stress in Ehrlich's ascites carcinoma cells: an insight into the mechanism of induction. *J.Venom.Anim Toxins.Incl.Trop.Dis.*, 19, 10.
- Alonso, D., Khalil, Z., Satkunanthan, N., & Livett, B. G. (2003). Drugs from the sea: conotoxins as drug leads for neuropathic pain and other neurological conditions. *Mini.Rev.Med.Chem.*, 3, 785-787.
- Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clin.Chem.*, 49, 1041-1044.
- Bandyopadhyay, P. K., Stevenson, B. J., Ownby, J. P., Cady, M. T., Watkins, M., & Olivera, B. M. (2008). The mitochondrial genome of Conus textile, coxI-coxII intergenic sequences and Conoidean evolution. *Mol.Phylogenet.Evol.*, 46, 215-223.
- Bias, D., Dutertre, S., Gerbault, A., Menou, J. L., Offord, R., Favreau, P. et al. (2009). Comparative proteomic study of the venom of the piscivorous cone snail Conus consors. *J.Proteomics.*, 72, 210-218.
- Biggs, J. S., Olivera, B. M., & Kantor, Y. I. (2008). Alpha-conopeptides specifically expressed in the salivary gland of Conus pulicarius. *Toxicon*, 52, 101-105.
- Bouchet, P., Rocroi, J. P., Fryda, J., Hausdorf, B., Ponder, W., Valdes, A. et al. (2005). Classification and nomenclator of gastropod families. *Malacologia*, 47, 1-368.
- Buczek, O., Bulaj, G., & Olivera, B. M. (2005). Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol.Life Sci.*, 62, 3067-3079.

- Colinge, J., Masselot, A., Giron, M., Dessingy, T., & Magnin, J. (2003). OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics.*, 3, 1454-1463.
- Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., & Fainzilber, M. (2001). Mechanisms for evolving hypervariability: the case of conopeptides. *Mol.Biol.Evol.*, 18, 120-131.
- Dave, K. & Lahiry, A. (2012). Conotoxins: Review and Docking Studies to determine potentials of Conotoxin as an Anticancer Drug Molecule. *Curr.Top.Med.Chem.*, 12, 845-851.
- Duda, T. F., Jr. & Palumbi, S. R. (1999). Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc.Natl.Acad.Sci.U.S.A*, 96, 6820-6823.
- Dutertre, S., Biass, D., Stocklin, R., & Favreau, P. (2010). Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon*, 55, 1453-1462.
- Dutertre, S., Jin, A. H., Kaas, Q., Jones, A., Alewood, P. F., & Lewis, R. J. (2013). Deep venomomics reveals the mechanism for expanded Peptide diversity in cone snail venom. *Mol.Cell Proteomics.*, 12, 312-329.
- Dutertre, S., Jin, A. H., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V. et al. (2014). Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat Commun*, 5, 3521.
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS.Comput.Biol.*, 7, e1002195.
- Favreau, P., Krimm, I., Le, G. F., Bobenrieth, M. J., Lamthanh, H., Bouet, F. et al. (1999). Biochemical characterization and nuclear magnetic resonance structure of novel

alpha-conotoxins isolated from the venom of *Conus consors*. *Biochemistry*, 38, 6317-6326.

Flecker, H. (1936). Cone Shell Mollusc Poisoning, With Report of a Fatal Case. *The Medical Journal of Australia*, 1, 464-466.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol.Biol.Evol.*, 27, 221-224.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst.Biol.*, 59, 307-321.

Han, Y., Huang, F., Jiang, H., Liu, L., Wang, Q., Wang, Y. et al. (2008). Purification and structural characterization of a D-amino acid-containing conopeptide, conomarphin, from *Conus marmoreus*. *FEBS J.*, 275, 1976-1987.

Hocking, H. G., Gerwig, G. J., Dutertre, S., Violette, A., Favreau, P., Stocklin, R. et al. (2013). Structure of the O-Glycosylated Conopeptide CcTx from *Conus consors* Venom. *Chemistry.*, 19, 870-879.

Hu, H., Bandyopadhyay, P. K., Olivera, B., & Yandell, M. (2011). Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC.Genomics*, 12, 60.

Jakubowski, J. A., Kelley, W. P., Sweedler, J. V., Gilly, W. F., & Schulz, J. R. (2005). Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J.Exp.Biol.*, 208, 2873-2883.

Jimenez, E. C., Olivera, B. M., & Cruz, L. J. (1983). Localization of enzymes and possible toxin precursors in granules from *Conus striatus* venom. *Toxicon*, 21, 199-202.

- Jin, A. H., Dutertre, S., Kaas, Q., Laverne, V., Kubala, P., Lewis, R. J. et al. (2013). Transcriptomic Messiness in the Venom Duct of *Conus miles* Contributes to Conotoxin Diversity. *Mol.Cell Proteomics.*, 12, 3824-3833.
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput.Appl.Biosci.*, 8, 275-282.
- Kaas, Q., Westermann, J. C., & Craik, D. J. (2010). Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon*, 55, 1491-1509.
- Kaas, Q., Westermann, J. C., Halai, R., Wang, C. K., & Craik, D. J. (2008). ConoServer, a database for conopeptide sequences and structures. *Bioinformatics.*, 24, 445-446.
- Kapono, C. A., Thapa, P., Cabalteja, C. C., Guendisch, D., Collier, A. C., & Bingham, J. P. (2013). Conotoxin truncation as a post-translational modification to increase the pharmacological diversity within the milked venom of *Conus magus*. *Toxicon*, 70, 170-178.
- Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol.Biol.Evol.*, 30, 772-780.
- Krais, A. M., Hautefeuille, A. H., Cros, M. P., Krutovskikh, V., Tournier, J. M., Birembaut, P. et al. (2011). CHRNA5 as negative regulator of nicotine signaling in normal and cancer bronchial cells: effects on motility, migration and p63 expression. *Carcinogenesis*, 32, 1388-1395.
- Le, G. F., Favreau, P., Benoit, E., Mattei, C., Bouet, F., Menou, J. L. et al. (1999). A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na⁺-dependent mechanism. *Eur.J.Neurosci.*, 11, 3134-3142.

- Leonardi, A., Biass, D., Kordis, D., Stocklin, R., Favreau, P., & Krizaj, I. (2012). Conus consors Snail Venom Proteomics Proposes Functions, Pathways, and Novel Families Involved in Its Venomic System. *J.Proteome.Res.*, 11, 5046-5058.
- Lewis, R. J., Dutertre, S., Vetter, I., & Christie, M. J. (2012). Conus venom Peptide pharmacology. *Pharmacol.Rev.*, 64, 259-298.
- Lluisma, A. O., Milash, B. A., Moore, B., Olivera, B. M., & Bandyopadhyay, P. K. (2012). Novel venom peptides from the cone snail Conus pulicarius discovered through next-generation sequencing of its venom duct transcriptome. *Mar.Genomics*, 5, 43-51.
- Luo, S., Christensen, S., Zhangsun, D., Wu, Y., Hu, Y., Zhu, X. et al. (2013). A Novel Inhibitor of $\alpha 9\alpha 10$ Nicotinic Acetylcholine Receptors from Conus vexillum Delineates a New Conotoxin Superfamily. *PLoS.One.*, 8, e54648.
- Marshall, J., Kelley, W. P., Rubakhin, S. S., Bingham, J. P., Sweedler, J. V., & Gilly, W. F. (2002). Anatomical correlates of venom production in Conus californicus. *Biol.Bull.*, 203, 27-41.
- Menez, A., Stocklin, R., & Mebs, D. (2006). 'Venomics' or : The venomous systems genome project. *Toxicon*, 47, 255-259.
- Miljanich, G. P. (2004). Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr.Med.Chem.*, 11, 3029-3040.
- Moller, C., Vanderweit, N., Bubis, J., & Mari, F. (2013). Comparative analysis of proteases in the injected and dissected venom of cone snail species. *Toxicon*, 65C, 59-67.
- Muttenthaler, M., Akondi, K. B., & Alewood, P. F. (2011). Structure-activity studies on α -conotoxins. *Curr.Pharm.Des*, 17, 4226-4241.

- Olivera, B. M., Gray, W. R., Zeikus, R., McIntosh, J. M., Varga, J., Rivier, J. et al. (1985). Peptide neurotoxins from fish-hunting cone snails. *Science*, 230, 1338-1343.
- Olivera, B. M., Walker, C., Cartier, G. E., Hooper, D., Santos, A. D., Schoenfeld, R. et al. (1999). Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann.N.Y.Acad.Sci.*, 870, 223-237.
- Pi, C., Liu, J., Peng, C., Liu, Y., Jiang, X., Zhao, Y. et al. (2006). Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics*, 88, 809-819.
- Puillandre, N., Bouchet, P., Duda, T. F., Jr., Kauferstein, S., Kohn, A. J., Olivera, B. M. et al. (2014). Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol.Phylogenet.Evol.*, 78C, 290-303.
- Puillandre, N., Koua, D., Favreau, P., Olivera, B. M., & Stocklin, R. (2012). Molecular phylogeny, classification and evolution of conopeptides. *J.Mol.Evol.*, 74, 297-309.
- Puillandre, N., Samadi, S., Boisselier, M. C., Sysoev, A. V., Kantor, Y. I., Cruaud, C. et al. (2008). Starting to unravel the toxoglossan knot: molecular phylogeny of the "turrids" (Neogastropoda: Conoidea). *Mol.Phylogenet.Evol.*, 47, 1122-1134.
- Puillandre, N., Watkins, M., & Olivera, B. M. (2010). Evolution of conus Peptide genes: duplication and positive selection in the α -superfamily. *J.Mol.Evol.*, 70, 190-202.
- Safavi-Hemami, H., Gorasia, D. G., Steiner, A. M., Williamson, N. A., Karas, J. A., Gajewiak, J. et al. (2012). Modulation of Conotoxin Structure and Function Is Achieved through a Multienzyme Complex in the Venom Glands of Cone Snails. *J.Biol.Chem.*, 287, 34288-34303.

- Safavi-Hemami, H., Siero, W. A., Gorasia, D. G., Young, N. D., Macmillan, D., Williamson, N. A. et al. (2011). Specialisation of the venom gland proteome in predatory cone snails reveals functional diversification of the conotoxin biosynthetic pathway. *J.Proteome.Res.*, 10, 3904-3919.
- Salisbury, S. M., Martin, G. G., Kier, W. M., & Schulz, J. R. (2010). Venom kinematics during prey capture in *Conus*: the biomechanics of a rapid injection system. *J.Exp.Biol.*, 213, 673-682.
- Taylor, J. D., Kantor, Y. I., & Sysoev, A. V. (1993). Foregut anatomy, feeding mechanisms, relationships and classification of the Conoidea (=Toxoglossa)(Gastropoda). *Bulletin of the Natural History Museum of London (Zoology)*, 59, 125-170.
- Teichert, R. W., Jacobsen, R., Terlau, H., Yoshikami, D., & Olivera, B. M. (2007). Discovery and characterization of the short kappaA-conotoxins: a novel subfamily of excitatory conotoxins. *Toxicon*, 49, 318-328.
- Terlau, H., Shon, K. J., Grilley, M., Stocker, M., Stuhmer, W., & Olivera, B. M. (1996). Strategy for rapid immobilization of prey by a fish-hunting marine snail. *Nature*, 381, 148-151.
- Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stocklin, R. et al. (2012). High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. *Toxicon*, 59, 34-46.
- Tucker, J. K. & Tenorio, M. J. (2009). *Systematic Classification of Fossil and Recent Conoidean Gastropods*. (- ed.) Hackenheim.
- Violette, A., Biass, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F. et al. (2012a). Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-

hunting cone snail using a combined proteomic and transcriptomic approach.

J.Proteomics., 75, 5215-5225.

Violette, A., Leonardi, A., Piquemal, D., Terrat, Y., Biass, D., Dutertre, S. et al. (2012b).

Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar.Drugs*, 10, 258-280.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview

Version 2--a multiple sequence alignment editor and analysis workbench.

Bioinformatics., 25, 1189-1191.

6.7 Concluding remarks

Not only does this study integrate the results of all previous studies carried out in the field of *C. consors* venom composition, but it also highlights the emergence of new venom components and thereby provides a broader and perhaps more definitive picture of the *C. consors* venom production and maturation processes.

This time the identification covered about 28% of the detected masses and again the combined transcriptomic and proteomic data strategy proved instrumental in rapid identification of hundreds of components, 299 peptides in this case with the dissected venom. Complex PTMs were revealed such as O-glycosylation. In parallel with new sequences found in different superfamilies – such as A-, M-, O-, S- and T-superfamilies – new families of conopeptides, bearing no disulfide bridges, were found in the T-superfamily. As already suggested by both the injectable venom study and another study¹⁹, the hyper-variability of C- and N- termini in certain conopeptide isoforms is a quite interesting phenomenon that results from the cone snail's need to rapidly evolve and modify its venom cocktail for optimal survival.

6.8 Reference List

1. Miller, J. R.; Koren, S.; Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* 2010, 95 (6), 315-327.
2. Kumar, S.; Blaxter, M. L. Comparing de novo assemblers for 454 transcriptome data. *BMC. Genomics* 2010, 11, 571.
3. Koua, D. Bioinformatics tools to assist drug candidate discovery in venom gland transcriptomes. University of Geneva, PhD Thesis, No. Sc. 4471, 2012.
4. Koua, D.; Brauer, A.; Laht, S.; Kaplinski, L.; Favreau, P.; Remm, M.; Lisacek, F.; Stocklin, R. ConoDICTor: a tool for prediction of conopeptide superfamilies. *Nucleic Acids Res.* 2012, 40 (Web Server issue), W238-W241.
5. Koua, D.; Laht, S.; Kaplinski, L.; Stocklin, R.; Remm, M.; Favreau, P.; Lisacek, F. Position-specific scoring matrix and hidden Markov model complement each other for the prediction of conopeptide superfamilies. *Biochim. Biophys. Acta* 2013, 1834 (4), 717-724.
6. Laht, S.; Koua, D.; Kaplinski, L.; Lisacek, F.; Stocklin, R.; Remm, M. Identification and classification of conopeptides using profile Hidden Markov Models. *Biochim. Biophys. Acta* 2011, 1824 (3), 488-492.
7. Tayo, L. L.; Lu, B.; Cruz, L. J.; Yates, J. R., III Proteomic analysis provides insights on venom processing in *Conus textile*. *J. Proteome. Res.* 2010, 9 (5), 2292-2301.
8. Buczek, O.; Bulaj, G.; Olivera, B. M. Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol. Life Sci.* 2005, 62 (24), 3067-3079.
9. Jimenez, E. C.; Craig, A. G.; Watkins, M.; Hillyard, D. R.; Gray, W. R.; Gulyas, J.; Rivier, J. E.; Cruz, L. J.; Olivera, B. M. Bromocontryphan: post-translational bromination of tryptophan. *Biochemistry* 1997, 36 (5), 989-994.
10. Craig, A. G.; Jimenez, E. C.; Dykert, J.; Nielsen, D. B.; Gulyas, J.; Abogadie, F. C.; Porter, J.; Rivier, J. E.; Cruz, L. J.; Olivera, B. M.; McIntosh, J. M. A novel post-translational modification involving bromination of tryptophan. Identification of the residue, L-6-

bromotryptophan, in peptides from *Conus imperialis* and *Conus radiatus* venom. *J. Biol. Chem.* 1997, 272 (8), 4689-4698.

11. Rigby, A. C.; Lucas-Meunier, E.; Kalume, D. E.; Czerwiec, E.; Hambe, B.; Dahlqvist, I.; Fossier, P.; Baux, G.; Roepstorff, P.; Baleja, J. D.; Furie, B. C.; Furie, B.; Stenflo, J. A conotoxin from *Conus textile* with unusual posttranslational modifications reduces presynaptic Ca^{2+} influx. *Proc. Natl. Acad. Sci. U. S. A* 1999, 96 (10), 5758-5763.

12. Ahrne, E.; Muller, M.; Lisacek, F. Unrestricted identification of modified proteins using MS/MS. *Proteomics*. 2010, 10 (4), 671-686.

13. Yates, J. R., III; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* 1995, 67 (8), 1426-1436.

14. Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004, 20 (9), 1466-1467.

15. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20 (18), 3551-3567.

16. Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T.; Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics*. 2003, 3 (8), 1454-1463.

17. Le, G. F.; Favreau, P.; Benoit, E.; Mattei, C.; Bouet, F.; Menou, J. L.; Menez, A.; Letourneux, Y.; Molgo, J. A new conotoxin isolated from *Conus consors* venom acting selectively on axons and motor nerve terminals through a Na^{+} -dependent mechanism. *Eur. J. Neurosci.* 1999, 11 (9), 3134-3142.

18. Jimenez, E. C.; Olivera, B. M.; Cruz, L. J. Localization of enzymes and possible toxin precursors in granules from *Conus striatus* venom. *Toxicon* 1983, 21 (Suppl.3), 199-202.

19. Dutertre, S.; Jin, A. H.; Kaas, Q.; Jones, A.; Alewood, P. F.; Lewis, R. J. Deep venomomics reveals the mechanism for expanded Peptide diversity in cone snail venom. *Mol. Cell Proteomics*. 2013, 12 (2), 312-329.

7. Concluding remarks and perspectives

7.1 Venom peptide and protein content: towards fully integrated venomics approach.

The study, on which this thesis is based, mainly focused on the sequence characterisation of the venom content of a cone snail and commanded the development and optimisation of methods and techniques required to extract a maximum of peptide or protein sequence information. The implementation of a fully integrated venomics approach involving the combination of transcriptomic data on the one hand and peptidomic and proteomic data on the other hand proved quite judicious to achieve this task, in a reasonable amount of time. Many cone snail venomics studies have since adopted this approach to contribute to the characterisation of many more new venom components and enrich the venom sequence databases¹⁻³. In addition, the contemporary development of dedicated bioinformatic platforms, such as TATools⁴, helped to manage, analyse and cross-link the growing amount of data generated by peptidomics and transcriptomics, paving the way to unlimited future applications.

7.2 Biological mechanisms in the venom duct.

In spite of the new insights brought to the biological understanding of the venom production and maturation through the characterisation of new venom compounds, the exact mechanisms that occur in the venom duct – from production to envenomation – are still not quite fully understood. One point that still remains unclear is how a cone snail can present virtually totally different injected venom profiles from one situation to another⁵. A recent study

attempted to bring an answer to that question by proposing the hypothesis that an initial stimulus (predatory or defensive) is perceived by the cone snail via its sensors to activate one of two separate neuronal circuits. One, a predation-triggered stimulus, activates the distal venom duct portion, causing the release of predatory venom peptides into the venom duct lumen; the other, a defence-prompted stimulus, activates the proximal venom duct part, causing the release of defensive venom peptides into the lumen (see Figure 44). The contraction of the muscular venom bulb then moves the lumen content to the proboscis to generate either a predation- or a defence-prompted venom profile.

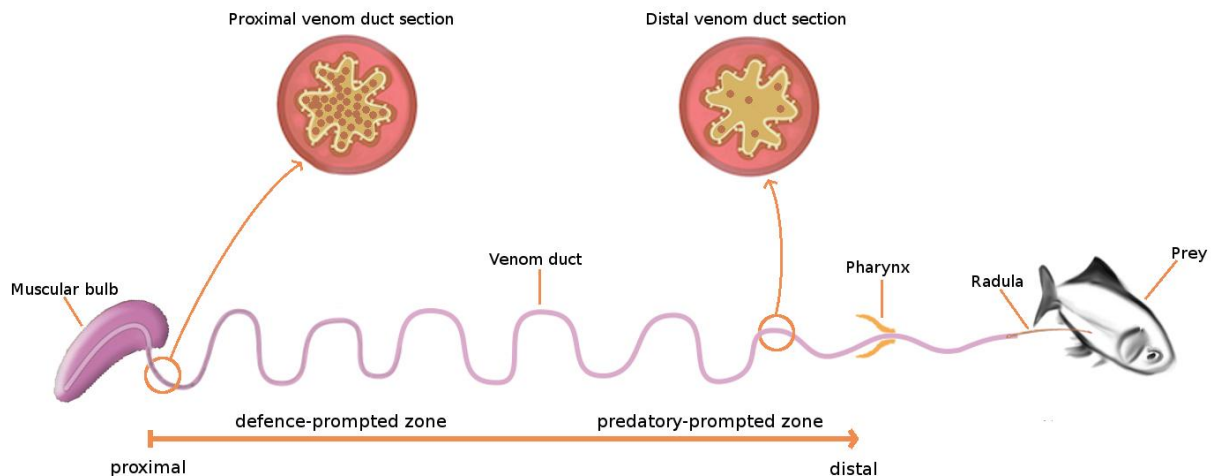


Figure 44: Simplified scheme of the venom apparatus.

This hypothesis is quite interesting, but still draws some questions on the time necessary to release the two different venom profiles after stimulus and how the venom bulb allows a rapid switch between the predation- and defence-prompted venom profiles. The results of the *Conus consors* dissected venom study and earlier studies on cone snail venom granules strongly suggest that these granules and their possible regulation may play an important role. Indeed, the way the venom granules are distributed along the venom duct may suggest that defence-prompted peptides are packed in the venom granules whereas the predatory-

prompted peptides are directly released in the venom lumen for immediate use. This would entail that, when a defence stimulus is received, the cone snail strongly contracts its venom bulb and thereby enables it to inject the venom granules containing the defence-prompted peptides already present in the lumen. These studies show the importance of different sampling methods to understand the biological mechanisms during venom processing and delivery.

7.3 Applications: pharmacological tools and drug candidates

Beyond the biological aspects and the sequence data gathered on the peptidomic, transcriptomic and genomic (genome to be published) levels, the integrated venomomics approach of the CONCO project has led to the direct application of some of the *C. consors* venom peptides.

The Xep-018, a μ -conopeptide (CnIIIC) inhibitor of the neuromuscular transmission, preferentially on $\text{Na}_v1.4$ receptor⁶, is still undergoing pre-clinical studies in view of its development into an anti-pain/anaesthetic myorelaxant drug. The same molecule has already been successfully released on the cosmetics market as instant line relaxer (<http://www.activen.ch/?page=xep>).

A new conopeptide called tau-CnVA was identified thanks to the peptidomic/transcriptomic strategy⁷. This peptide was chemically synthesized and its 3D-structure solved by NMR. It was found to interact selectively with the somatostatine sst3 receptor, which currently makes it the only known toxin peptide to interact with this GPCR subfamily.

Different δ -conopeptides have been discovered through this venomomics approach. This conopeptide family appears to be quite specific to piscivorous cone snails, with a few exceptions regarding molluscivorous cone snails. Their structure and physico-chemical properties have made them very difficult to study and to synthesise. Using a new synthesis

approach, three δ -conopeptides and two analogs were synthesised, leading to the determination of Na_v subtype selectivity⁸. This new method, applicable to other peptides with comparable complex structures, would enable new synthetic compounds as pharmacological tools to be provided to further understand Na_v channels – a very “hot topic” since the Na_v1.7 became one of the main targets for pain relief drug candidates⁹.

7.4 Perspectives for venomomics and drug candidate discovery

7.4.1 LC and MS instruments

In parallel to the studies presented here, different strategies were also developed in an attempt to improve the recovery rate of identified sequences over the total number of masses detected in a complex mixture like cone snail venoms.

Regarding the matter of liquid chromatography, the development of ultra high performance liquid chromatography (UHPLC) has greatly improved the efficiency of the chromatographic separations over standard HPLC and very high resolution separations are now obtained with natural products in a reasonable amount of time. This chromatographic technique was actually tested on *C. consors* dissected venom and proved that better separation and high resolution could be obtained without compromising the sample¹⁰. Turning to MS, major advances in the profiling of complex natural mixtures have also been achieved over the past few years. Nowadays, most of the QTOF instruments provide high resolution and high mass accuracy. However, in combination with UHPLC systems that provide very thin peaks in the chromatographic dimension, QTOF-MS instruments would require additional acquisition rate improvements to allow them to yield a greater number of MS and MS/MS spectra per chromatographic peak, which is a *sine qua non* to obtain good quality spectra for downstream analysis with automated deconvolution and peptide mass matching programmes. Recent MS instruments, using orbitrap or ion mobility spectrometry technology,

are now also well suited for such venomomics applications in combination with high resolution separation chromatography. Preliminary tests on *C. consors* venom samples have been made with such instruments for venomomics application, however the downstream MS/MS data processing and sorting with peptide mass matching programmes still need to be improved. In addition, the cost of such instruments still hinders their broad application in the venomomics field.

7.4.2 Improving PTM detection with new algorithms

In the venomomics strategy, improvements in sequencing efficiency inevitably run in parallel with the enhancement of PTM detection directly from MS/MS spectra and/or the integration of novel algorithms in peptide mass matching programmes for the correct assignment of PTM. Various cone snail venomomics studies have shown the importance of PTMs in conopeptides.

For example, a small in-house script was developed as part of this study to help, in a semi-automatic way, with the characterisation of glycosylated conopeptides or conoproteins. Not only was this script successfully used to properly interpret MS/MS spectra presenting sugar unit signals and characterise new sequences in the *C. consors* dissected venom study presented here, it also was implemented for the partial characterisation of a novel glycosyl hydrolase protein, Conohyal-Cn1, also discovered during the CONCO project¹¹.

Unfortunately, directly implementing such complex PTMs in peptide mass matching programmes will lead to an increasing number of false positives¹². A solution to the increasing amount of MS/MS data generated by venomomics projects might be to pre-process the data to search for known PTMs and send only a filtered version of the MS/MS data to the peptide mass matching programmes and use new algorithms to work directly by *de novo* sequencing to maximise the number of novel characterised sequences¹³.

7.4.3 Towards the direct implementation of new bio-assay technologies in the venomomics strategy

Finally, one can expect a direct integration of bio-assay techniques in the venomomics strategy in the future to further enhance the speed of drug discovery. With the increasing development of chip-based nano-assays, the gap with even more integrated projects is narrowing. Recently, a study using in parallel chip-based bioassay and MS coupled to nano-liquid chromatography demonstrated the feasibility of detecting cone snail venom compound masses and, in parallel, screening for their activity on acetylcholine binding proteins, a mimic of the nAChR¹⁴. Another option would be the implementation of cell-specific constellation assay in the same manner as mentioned above¹⁵. In this context, the combination of latest generation MS instruments and microfluidic chip-based bioassays would enable one to perform both the screening of a compound for a given biological activity and its direct sequence characterisation via MS/MS and thereby speed up drug discovery.

7.5 Reference list

1. Dutertre, S.; Jin, A. H.; Kaas, Q.; Jones, A.; Alewood, P. F.; Lewis, R. J. Deep venomomics reveals the mechanism for expanded Peptide diversity in cone snail venom. *Mol. Cell Proteomics*. **2013**, *12* (2), 312-329.
2. Prashanth, J. R.; Lewis, R. J.; Dutertre, S. Towards an integrated venomomics approach for accelerated conopeptide discovery. *Toxicon* **2012**, *60* (4), 470-477.
3. Prashanth, J. R.; Brust, A.; Jin, A. H.; Alewood, P. F.; Dutertre, S.; Lewis, R. J. Cone snail venomomics: from novel biology to novel therapeutics. *Future. Med. Chem.* **2014**, *6* (15), 1659-1675.
4. Koua, D. Bioinformatics tools to assist drug candidate discovery in venom gland transcriptomes. University of Geneva, PhD Thesis, No. Sc. 4471, 2012.
5. Dutertre, S.; Biass, D.; Stocklin, R.; Favreau, P. Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* **2010**, *55* (8), 1453-1462.
6. Favreau, P.; Benoit, E.; Hocking, H. G.; Carlier, L.; D', H. D.; Leipold, E.; Markgraf, R.; Schlumberger, S.; Cordova, M. A.; Gaertner, H.; Paolini-Bertrand, M.; Hartley, O.; Tytgat, J.; Heinemann, S. H.; Bertrand, D.; Boelens, R.; Stocklin, R.; Molgo, J. A novel micro-conopeptide, CnIIIC, exerts potent and preferential inhibition of Na(V) 1.2/1.4 channels and blocks neuronal nicotinic acetylcholine receptors. *Br. J. Pharmacol.* **2012**, *166* (5), 1654-1668.
7. Petrel, C.; Hocking, H. G.; Reynaud, M.; Upert, G.; Favreau, P.; Biass, D.; Paolini-Bertrand, M.; Peigneur, S.; Tytgat, J.; Gilles, N.; Hartley, O.; Boelens, R.; Stocklin, R.; Servent, D. Identification, structural and pharmacological characterization of tau-CnVA, a conopeptide that selectively interacts with somatostatin sst3 receptor. *Biochem. Pharmacol.* **2013**, *85* (11), 1663-1671.
8. Peigneur, S.; Paolini-Bertrand, M.; Gaertner, H.; Biass, D.; Violette, A.; Stocklin, R.; Favreau, P.; Tytgat, J.; Hartley, O. delta-Conotoxins Synthesized Using an Acid-

cleavable Solubility Tag Approach Reveal Key Structural Determinants for NaV Subtype Selectivity. *J. Biol. Chem.* **2014**, 289 (51), 35341-35350.

9. Dib-Hajj, S. D.; Cummins, T. R.; Black, J. A.; Waxman, S. G. From genes to pain: Na_v 1.7 and human pain disorders. *Trends Neurosci.* **2007**, 30 (11), 555-563.
10. Eugster, P. J.; Biass, D.; Guilleme, D.; Favreau, P.; Stocklin, R.; Wolfender, J. L. Peak capacity optimisation for high resolution peptide profiling in complex mixtures by liquid chromatography coupled to time-of-flight mass spectrometry: Application to the Conus consors cone snail venom. *J. Chromatogr. A* **2012**, 1259, 187-199.
11. Violette, A.; Leonardi, A.; Piquemal, D.; Terrat, Y.; Biass, D.; Dutertre, S.; Noguier, F.; Ducancel, F.; Stocklin, R.; Krizaj, I.; Favreau, P. Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of conus venoms. *Mar. Drugs* **2012**, 10 (2), 258-280.
12. Ahrne, E.; Muller, M.; Lisacek, F. Unrestricted identification of modified proteins using MS/MS. *Proteomics*. **2010**, 10 (4), 671-686.
13. Bhatia, S.; Kil, Y. J.; Ueberheide, B.; Chait, B. T.; Tayo, L.; Cruz, L.; Lu, B.; Yates, J. R., III; Bern, M. Constrained de novo sequencing of conotoxins. *J. Proteome. Res.* **2012**, 11 (8), 4191-4200.
14. Heus, F.; Otvos, R. A.; Aspers, R. L.; van, E. R.; Halff, J. I.; Ehlers, A. W.; Dutertre, S.; Lewis, R. J.; Wijmenga, S.; Smit, A. B.; Niessen, W. M.; Kool, J. Miniaturized bioaffinity assessment coupled to mass spectrometry for guided purification of bioactives from toad and cone snail. *Biology. (Basel)* **2014**, 3 (1), 139-156.
15. Imperial, J. S.; Cabang, A. B.; Song, J.; Raghuraman, S.; Gajewiak, J.; Watkins, M.; Showers-Corneli, P.; Fedosov, A.; Concepcion, G. P.; Terlau, H.; Teichert, R. W.; Olivera, B. M. A family of excitatory peptide toxins from venomous crassispirine snails: using Constellation Pharmacology to assess bioactivity. *Toxicon* **2014**, 89, 45-54.

Acknowledgements

I would like to thank the members of the jury who accepted to surrender part of their time to read and evaluate the work presented herewith, namely:

Prof. Eric Allémann from the University of Geneva, President of the jury,

Prof. Juan J. Calvete from the Instituto de Biomedicina de Valencia, external expert,

Prof. Jean-Luc Wolfender from the University of Geneva, internal expert,

Dr. Frédérique Lisacek from the Swiss Institute of Bioinformatics, internal expert,

Prof. Denis Hochstrasser from the University of Geneva, Co-director,

Dr. Reto Stöcklin from the Atheris Laboratories, Co-director.

I would like to thank all those who, believing in the project, provided support and help, thus allowing the achievement of this thesis.

These include Prof. Denis Hochstrasser and Dr. Reto Stöcklin for giving me the opportunity to accomplish this thesis and for being my co-directors. I would like to thank Dr. Reto Stöcklin for welcoming me in his company and laboratories to achieve my work and thereby giving me the opportunity to collaborate on many of his company's other projects.

I would like to pay tribute to all the partners of the CONCO project for the very fine scientific collaborations and the equally fine human and scientific experiences generated all along these years –the scientific collection mission in French Polynesia, for example, will remain etched in my memory forever.

Turning to the University of Geneva, I would like to thank Prof. Jean-Luc Wolfender and his team, especially Dr. Philippe Eugster, as well as Prof. Oliver Hartley and his team for their very kind collaboration.

I would like to thank Dr. Frédérique Lisacek from the Swiss Institute of Bioinformatics for her invaluable help in the final straight line of the manuscript writing race.

I would like to thank all my Atheris Laboratories colleagues for their daily help and advice, and for all the good times we had together, namely: Philippe (Philou), Sophie, Laure, Fred, Célia, Olivier, Florence, Florence, Roman, Greg, Guillaume, Dominique, Vera, Coralie, Maya, Estelle, Lou, Christophe, Sébastien, Nathalie, Aude, Cécile, Nicolas, Francine, Sylvie, Alain and Xavier.

Last, but certainly not least, thanks must also go to all my friends and my family, my parents, my wife and my children for their support, kindness and patience.