REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author GENNITS, A-J

COPYRIGHT
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS
Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of UCL

☐ This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

Characterisation of *Conus* Venom Using Polypeptide Separation Methods and Mass Spectrometric Analysis

ALBERT J. GERRITS

DEPARTMENT OF BIOCHEMISTRY AND
MOLECULAR BIOLOGY
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy from the University of London

UMI Number: U592834

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U592834
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Each of the 500 known *Conus* species has its own distinctive, complex and peptide rich venom. Approximately 100 different peptide toxins, so called conotoxins can be expressed with little similarity between the species. An overwhelming majority of conotoxins are probably targeted selectively to specific ion channels. Because conotoxins can discriminate between closely related subtypes of ion channels, many of the characterised conotoxins are used as pharmacological agents in ion channel research. Several have direct diagnostic and therapeutic potential. Although the venoms have been studied for the last two decades, only a miniscule fraction of the postulated conotoxins have been identified.

The studies presented in this dissertation were set out with the aim to further characterise the relative unknown venom of *Conus ventricosus* and to explore the protein content of *Conus* venoms in general. In total five different venoms were investigated.

Chapter three describes the analysis of protein content in various *Conus* venoms. Although both one and two-dimensional gel electrophoresis was applied, no known toxin precursors were found. Proteins identified are mainly housekeeping proteins, such as actin. One of the strategies applied for finding novel conotoxins involved a multi dimensional liquid chromatography experiment where a trypsin digest was performed in between two separations before mass spectrometric analysis. One putative novel partial conotoxin was found.

Chapters five and six describe the mass spectrometric analysis of *Conus ventricosus* and *textile* venom after a two-dimensional liquid chromatography separation. With the methods applied, besides verifying a number of putative sequences, several novel conotoxins were found, in particular a number of so called contryphans. One of these contryphans found contained a number of post-translational modifications, amongst them hydroxyproline and the very rare brominated tryptophan.

Samenvatting voor de leek

Conus ventricosus en *Conus textile* behoren tot de diersoort *Conus*. *Conus* zijn zeeslakken die voorkomen in sub-tropische wateren over de gehele wereld. Deze zeeslakken voeden zich met vissen, wormen en andere zeeslakken. Om de prooien te verlammen en te doden, bezitten ze over complexe systemen om gif te injecteren in hun prooi. Dit gif bestaat uit een cocktail van kleine eiwitten, conotoxins genaamd. De cocktail is zeer toxisch; zo zijn er gevallen bekend van vergiftingen bij mensen met de dood tot gevolg.

Deze peptiden zijn over het algemeen inhibitoren van ion kanalen die ionen zoals natrium en kalium transporteren door een membraan. Dit transport van ionen is belangrijk voor de elektrische signalen die worden gegenereerd door zenuwen die aan de hersenen moeten worden doorgegeven. Door dit proces te verstoren, treden verlammingsverschijnselen op die, in het geval van kleine organismen, de dood tot gevolg hebben.

Wetenschappers zijn al decennia lang geinteresseerd in de werking van deze conotoxins, maar hebben slechts enkele van de ongeveer 50.000 voorspelde conotoxins ontdekt en gekarakteriseerd.

Met de experimenten die zijn uitgevoerd voor dit promotie onderzoek zijn een aantal nieuwe conotoxins gevonden en structureel gekarakteriseerd. Conotoxins zijn met behulp van chromatografie geïsoleerd en met massa spectrometrie gekarakteriseerd. Een aantal van de gevonden conotoxins behoren tot de subgroep contryphans. Dit zijn, met één disulfide brug gebonden peptiden met veel modifikaties. Een zeer zeldzame modifikatie is gebromeerde tryptofaan, deze werd, samen met gehydroxyleerde proline aangetroffen op een van de contryfaans.

Acknowledgements

The work presented in this thesis would not have been possible without the help of many people. This section is therefore an attempt to thank all of those who played a crucial role. I have to start with someone, so therefore I would first and foremost like to thank Prof. Al Burlingame, for providing me with the position at the Ludwig Institute for Cancer Research, UCL Branch and making the whole ordeal possible. But also for funding, it seemed endless trips to San Francisco to the Mass Spec facility there. Without that luxury this thesis would not have existed. The experiences (both scientific and leisure), results and friends I gained there are irreplaceable and unforgettable.

However, the solid base of my PhD work was always in the Bioanalytical Laboratory of the LICR under the supervision of my "Doktorvater" Prof. Rainer Cramer. I would like to thank him for his support, advice and patience, as I know that working with me can be a challenge at times. His expertise will be sadly missed at the LICR.

"The coldest winter I spend was a summer in San Francisco" is a quote (supposedly) by Mark Twain (1835-1910). Anyway, whoever said that was right.

I would like to thank all the people at the UCSF Mass Spec facility for their help, however a few special thanks are necessary. First of all Kirk, thank you for all your help with everything, not only conducting experiments with me but also for making me enthusiastic about climbing. Robert, many thanks for the discussion and time we spend in the endless 'meetings' in the pubs around the campus. Also both trips to Yosemite National Park were absolutely out of this world. Craig, thank you ever so much for your help with the de novo sequencing. David, John, Fexia, Xin, Kati, 'old-man' Jim, Peter, Mike, Kris, Ronde, Aenoch, Keith & Emily (in random order) many thanks!

All the venom samples where provided by Dr. Mike Fainzilber. I would like to thank him not only for that, but also for bringing me into contact with Al, as this is how the whole story started.

My time in London was and is truly unforgettable. At the LICR there are literally too many people to thank. You have all been of great help in one way or another. Prof. Mike Waterfield, Head of LICR is greatly acknowledged for his supervision and optimism. He is of great help to all the PhD students, not only scientific.

From the ex-Cramer lab I would first like to thank my friend Andrew. I am not going to specify what he has done and what he meant for me, as this would substantially add up to the printing costs of this dissertation. Andy, "cheers mate!!" My dear Pedro, thank you for your friendship, hospitality, advice, humour and for correcting this work, also "cheers mate!!" Richard, many things would have literally been impossible without your programming skills. Not only for that but also for the after work pub crawls, yet another, "cheers mate!!" For the rest of the gang, Malc, Steve, Denis, John T, Mark, John S, Akunna, thank you so much for all your help!! Musarat, thank you for the discussions and over seas chatting.

I would also like to thank David Creasy and Patrick Emery from Matrix Science for their collaboration on various old and novel Mascot products.

Hong-Lin, it was a pleasure sharing an office with you, Taiwan and their people has a total different meaning to me now. Darling Mariana, your friendship is one of the most important things I gained from my time in London. Lets????

Some thank you's to non work related people: Tony and Jan, Anwyn, Paul, Paul, Paul (little Paul, free Willy and Mandonkey respectively), De Hems and London in general, you were all great!!!

I am sure that I missed out on a couple of people, but not intentionally!!! Whom ever you are thank you!!

One person whom I impossibly can not forget to thank is my best man and dear friend Sjouke. Many people stood at the beginning of this PhD, but only through your excellent supervision, advice and unconditional friendship (Gimme Shelter) this was made possible. Not many people are genuinely nice, you deserve a very special "cheers mate!!".

The most important people are and have to be my lovely parents. Pa en Ma voor jullie heb ik de "samenvatting" geschreven. Desondanks de soms grote afstanden is ons contact alleen maar sterker geworden. Veel personen zijn belangrijk geweest voor mij in deze periode, maar jullie stonden aan het begin. Bedankt voor al jullie steun and liefde . . . . .(en Berenburg)

To Nina. Life and what one achieves become insignificant if you can't share it with someone very special. I am very glad that my social life was always more important than my professional one and that therefore I married the "*Love Of My Life*" during this time although she was too far away for too long. You are special, you are my everything, Ja Volim Tebe, do neba . . . . . .

Abbreviations

The abbreviations that are used throughout this thesis are listed below. The abbreviations are subdivided in four categories: General; Chemicals; Mass spectrometry; Instrumentation and techniques. The meaning of the abbreviation is also given the first time used in the main text. Abbreviations of amino acids can be found in appendix 1.

General

Da	Dalton
LASER	Light Amplification by Stimulated Emission of Radiation
Mw	Molecular Weight
mmu	millimass units
pI	Isoelectric point
ppm	parts per million
RPM	Revolution per Minute
RT	Room Temperature
UV	Ultra Violet

Chemicals

ABC	Ammonium bicarbonate – NH_4HCO_3
ACN	Acetonitrile
ACTH	Adrenocorticotropic Hormone
APS	Ammonium Persulphate
BPB	Bromophenol Blue
C18	Octadecyl silane packing material
CCB	Colloidal Coomassie Blue
CHAPS	(3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate)
CHCA	Alpha cyano 4-hydroxy-cinnamic acid
Cy2	3-(4-carboxymethyl)phenylmethyl)-3'-ethyloxacarbocyanine halide N-hydroxy-succinimidyl ester
Cy3	1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxy-succinimidyl ester
Cy5	1-(5-carboxypentyl)-1'-methylindodicarbocyanine halide N-hydroxy-succinimidyl
DHB	2,5-dihydroxybenzoic acid
DTT	Dithiothreitol
FA	Formic Acid
IAM	Iodoacetamide
SDS	Sodium Dodecyl Sulphate
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic Acid
Tris	Tris-(hydroxymethyl)aminomethane

Mass spectrometry

CI	Chemical Ionization
CID	Collision Induced Dissociation
DE	Delayed Extraction
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FT	Fourier Tranformation
ICR	Ion Cyclotron Resonance
MALDI	Matrix-Assisted Laser Desorption/Ionization
$[MH]^+$	$[M+H]^+$ - Singly charged protonated ion
$[M2H]^{2+}$	$[M+2H]^{2+}$ - Doubly charged protonated ion
m/z	Mass to charge ratio
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
OaTOF	Orthogonal Acceleration Time Of Flight
PSD	Post-source Decay
Q-TOF	Quadrupole Orthogonal Acceleration Time Of Flight
TIC	Total Ion Count
TOF	Time Of Flight

Instrumentation and techniques

1DE	One-Dimensional Gel Electrophoresis
2DE	Two-Dimensional Gel Electrophoresis
2D-LC	Two-Dimensional Liquid Chromatography
BVA	Biological Variation Analysis
DIA	Difference In-gel Analysis
DIGE	2D-difference gel electrophoresis
HPLC	High Performance Liquid Chromatography
IEF	Iso-Electric Focussing
IPG	Immobilised pH Gradient
LC	Liquid Chromatography
LC-MS	Liquid Chromatography coupled online with Mass Spectromety
PAGE	Polyacrylamide Gel Electrophoresis
RP	Reversed Phase

Contents

Figures

Tables

1 Introduction

Marine snails from the genus *Conus* are venomous molluscs found in (sub-)tropical waters. These snails use small, structured peptide toxins (conotoxins) for prey capture, defense and competitor deterrence (Terlau and Olivera, 2004). Each of the known 500 *Conus* species can express approximately 100 different conotoxins, with little overlap between species (Röckel et al., 1995). Conotoxins that have been characterised so far display incredible specificity towards ion channels. Because of this specificity and the ability to discriminate between closely related subtypes of ion channels, these conotoxins are widely used as pharmacological agents in ion channel research, and have direct diagnostic and therapeutic potential. It is thought that an overwhelming majority of the predicted 50,000 conotoxins are targeted to specific ion channels.

1.1 An overview of Cone snails and conotoxins

Cone shells are found in all habitats from shallow to moderate depths mainly in (sub)-tropical oceans around the world, although some smaller species have been found on the southern coast of Australia [*], where the water temperature is only 5 – 12 °C. The sand-dwelling species possess very clean shells while those that live on top of sand or among rocks or rubble have a layer called a periostracum. Figure 1.1.1 shows the images of shells of two *Conus* species, namely *Conus ventricosus* and *Conus textile.*

1.1.1 *Conus* anatomy and life style

Cone shells are mainly nocturnal or crepuscular (i.e. they are active at twilight or just before dawn). Most *Conus* species (>150) prey on polychaete worms (vermivorous), some 70 feed upon gastropods including *Conus* (molluscivorous) while approximately 70 species prey on fish (piscivorous) (Jones and Bulaj, 2000). Some other *Conus* species prey on hemichordates. Hemichordates are distinguished by a tripartite (threefold) division of the body (Olivera, 1997). To make up for the lack of speed compared to their swift preys, especially the piscivorous species, these gastropods have developed an effective mechanism for immobilizing their prey. The piscivorous species can be divided into two groups, namely the "hook-and-line" fishing snails, which use a tooth like mini harpoon. And the "net-

[*] Source: http://grimwade.biochem.unimelb.edu.au/cone/index1.html

fishing" cone snails. An image of the latter is included in the bottom panel of Figure 1.1.2. With this technique, fish or a small school of fish are caught by extending its probosis like a net to catch multiple fish. In both cases the solution to the snail's lack of physical agility has been the development of a highly potent mixture of neuronal toxins which it uses to paralyze its prey.

Figure 1.1.1 Images of *Conus* shells.

Images of two *Conus* species are shown in this figure. In the left hand panel that of *Conus ventricosus* and in the right hand panel *Conus textile*. *Conus ventricosus* is found in the Mediterranean Sea and is classified as a molluscivorous species, while the piscivorous *Conus textile* is found in the Indo-Pacific Ocean. Images are courtesy of Eddie Hardy (http://www.gastropods.com).

To paralyze and capture their prey, this genus has developed a specialised venom apparatus. A modified radular tooth, which is used to inject the venom into the prey, functions as a disposable hypodermic needle (see Figure 1.1.2). The actual production and transfer of venom to the prey involves 3 broad steps: (i) synthesis, processing and packaging of peptide toxins. (ii) Generation and storage of radular teeth and transfer of a tooth to the tip of the proboscis. (iii) The final insertion of the tooth and injection of venom (Marshall et al., 2002). All these processes are operated by the venom apparatus in combination with the front part of the digestive tract. See Figure 1.1.2 for a schematic of the venom apparatus.

(i) The venom is produced in a long venom duct, often several times the length of the snail itself. The venom is pushed through the duct towards the proboscis by the muscular bulb, which is located at the posterior end of the tubular duct.

However, the sites of synthesis, and where the peptides are post-translationally modified remain unclear.

(ii) The teeth that are used for the injection of venom are made in the long arm of the radular sac. The completed teeth are stored in the short arm of the radular sac. A single tooth is transported to the pharynx. Finally the proboscis extends with the harpoon in the tip.

(iii) The current hypothesis is that the muscular bulb provides the force necessary to pierce the needle into the prey. This is done by forcefully pumping the venom into the needle, which is enclosed by the proboscis (Kohn and Hunter, 2001). However the exact mechanism is unclear.

Conus venom peptides were first isolated and characterised by Cruz *et al.* more then 20 years ago. Each of the approximate 500 species has its own unique and complex venom (Röckel et al., 1995). It contains an array of neurotoxins, so called conotoxins or conopeptides, that block a wide spectrum of ion-channels and neuronal receptors (Jones and Bulaj, 2000). The venoms contain up to 50 - 200 different peptides (Olivera, 1997), potentially leading to over 50,000 biologically active peptides. To date approximate 500 conotoxin sequences are published[†]. However many of the sequences are predicted sequences from cDNA libraries and only a few have been studied in detail.

[†] Source: http://au.expasy.org/cgi-bin/sprot-search-de?conotoxin

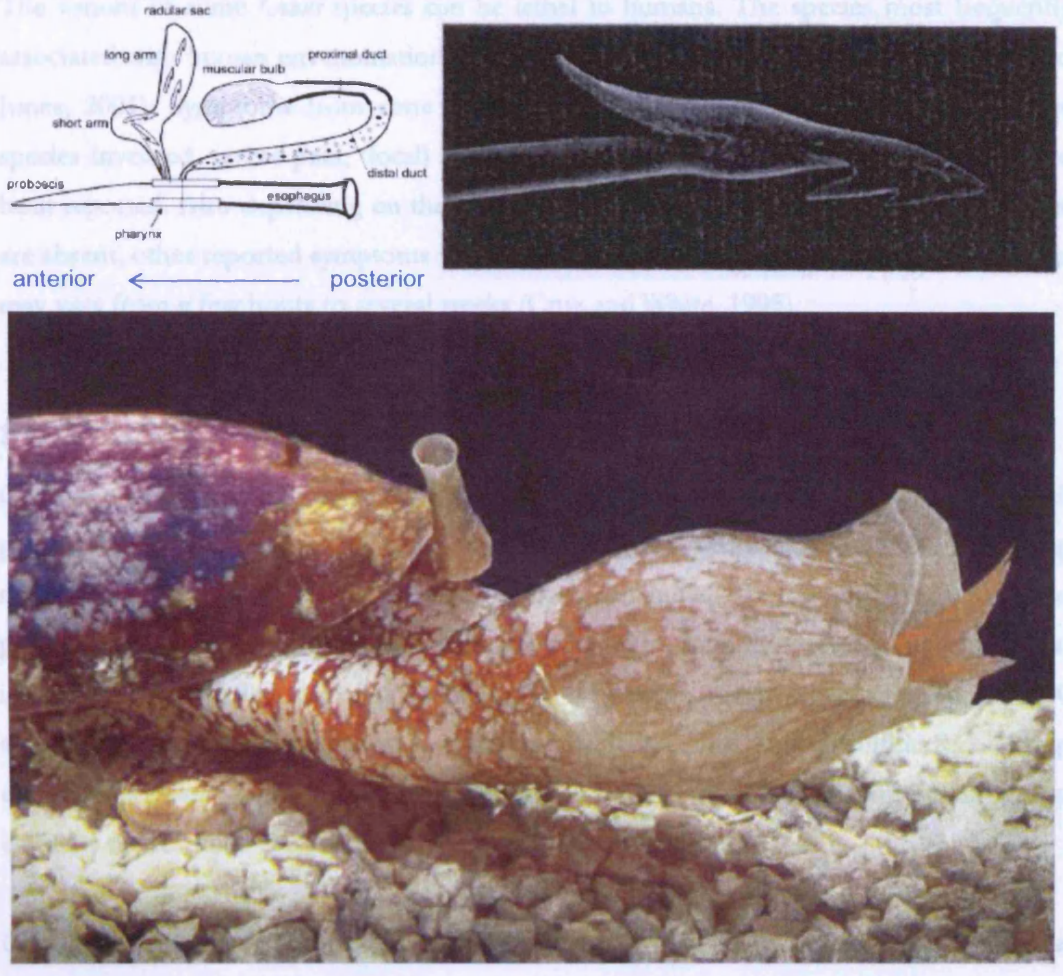


Figure 1.1.2 Anatomy and the catch . . .

The top left of the figure shows a schematic of the venom apparatus of *Conus* snails. 'Disposable hypodermic needle' are generated and stored in the radular sac. A scanning electron micrograph image of, the tip of the harpoon-like tooth of *Conus obscurus* is shown on the top right of the figure. The barbed, hollow tooth is used for injecting venom into the fish prey. One needle at the time is moved to the proboscis. It is thought that the force with which the muscular bulb is compressed, the venom is forced into the needle which then dislodges and is 'shot' towards the prey like a mini harpoon. The bottom of the figure shows an image of the capture of a fish by *Conus geographus* (The tip of the tail is just visible at the right hand side). *Conus geographus* is one of the most deadly species of *Conus*. The `Geography Cone' uses its distended rostrum as a net to capture fish. Multiple fish may be engulfed simultaneously and then individually envenomated. The figure on the left is adapted from Marshall et al. (Marshall et al., 2002), and the right hand side figure is taken from McIntosh et al. (McIntosh and Jones, 2001).

The venom of some *Conus* species can be lethal to humans. The species most frequently associated with human envenomations are *Conus textile* and *Conus geographus* (McIntosh and Jones, 2001). Symptoms from cone snail stings vary to some degree depending on the species involved, severe pain, (local) paralysis, difficulty in breathing and swallowing have been reported. Also depending on the severity of the envenomation, gag and other reflexes are absent, other reported symptoms were blurred and double vision. Recovery time, if any, may vary from a few hours to several weeks (Cruz and White, 1995).

1.1.2 *Conus* venom, the case of conotoxins

Components present in *Conus* derived venom are mainly made up of proteins and small peptides. These peptides are the active components when it comes to paralysing and killing their prey. The peptides found in *Conus* venom can initially be divided into two broad groups, namely non-disulfide rich and disulfide rich. Disulfide bridges are the most important post-translational modifications. The otherwise small peptides, typically 10 to 30 amino acids long, are shaped into rigid globular peptides. The 'Non-Disulfide-Rich' group contains those peptides with one or no disulfide bonds, while the 'Disulfide-Rich' group contains the so called conotoxins with two and more disulfides.

Conotoxins are named according to a systematic nomenclature, first proposed by Olivera *et al.* (Cruz et al., 1985) and later reviewed and updated by the same group (Gray et al., 1988). More recently it was updated by McIntosh and Olivera (McIntosh et al., 1999). This system of nomenclature was designed to provide information about the mechanisms of action, disulfide connectivity and species from which the toxin is derived. Some examples of conotoxin names are given in Figure 1.1.3.

μ - GIIIB ω - MVIIA δ - TxVIA

Figure 1.1.3 Examples of conotoxin nomenclature.

This figure shows a number of examples of conotoxin nomenclature. From left to right: mu conotoxin IIIB from *Conus geographus*, omega conotoxin VIIA from *Conus magnus* and delta conotoxin VIA from *Conus textile*. The middle peptide is also known as SNX-111 or Ziconotide and is currently under investigation in a stage III medical trial by Elan Pharmaceuticals (CA, USA).

The first letter (Greek alphabet) identifies the family to which the toxin belongs. For instance toxins in the sub-group of the M-superfamily that target sodium channels are classified as a mu-conotoxin. The second set of letters, or letter indicate the species from which the toxin was isolated (G = *geographus*; M = *magnus*; Tx = *textile*). The roman numerals indicate the disulfide connectivity. In general these refer to the superfamilies, I and II to the A-superfamily, III to the M-superfamily, IV refers to the T-superfamily, VI and VII denote the O-superfamily and VIII the S-superfamily. Figure 1.1.4 shows an overview of the classifications of conotoxins (Jones and Bulaj, 2000).

For peptides that have not been pharmacologically characterised, Olivera *et al.* developed a comparable naming strategy (Olivera et al., 1985), namely that all letters are substituted by lowercase letters and roman numerals are replaced by conventional letters. For example μ-GIIIB would be named as g3b. The Greek letter would, at this point is not assigned, as the molecular target is still unknown.

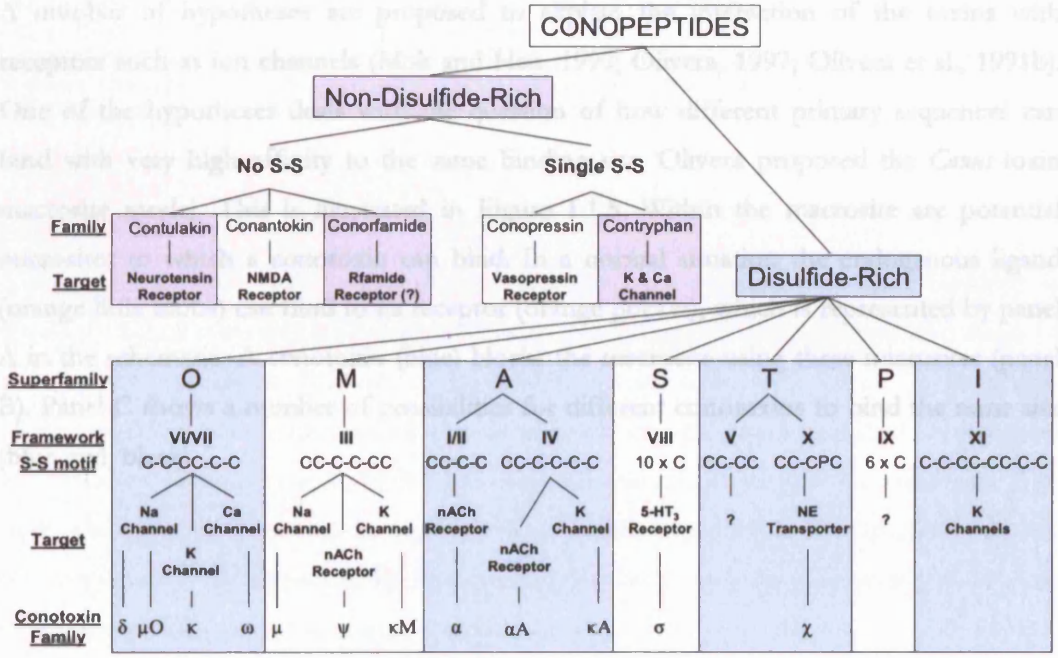


Figure 1.1.4 An overview of conotoxins

Peptides from *Conus* derived venom are divided into two broad groups, namely non-disulfide rich and disulfide rich. Members of each superfamily of conotoxins (multiple disulfide bridges) have a common highly conserved signal sequence in their precursors. The family members within the superfamily have a distinctive disulfide arrangement in the mature part of the toxin. The penultimate conotoxin family is defined by the target of the specific toxin. This figure is taken from Terlau and Olivera (Terlau and Olivera, 2004)

The importance of the strict disulfide arrangement to the survival of the species became clear after a study by Conticello *et al.* (Conticello et al., 2000). In this study 53 conotoxin

sequences derived from a cDNA library, were aligned. Not surprisingly the cysteines of the mature toxins show a high degree of conservation as did the signal peptides of the entire toxins. However, when examining the codons used for encoding the cysteine, something remarkable was revealed. It turns out that *Conus* have developed a mechanism where, the normal hypervariable conotoxins, there is specific codon preservation in specific regions of the toxin. For instance, the first cysteine in a CxxCxxCCxxCxxC scaffold was for 100% encoded by TGC. Cysteine can be encoded by two codons, namely TGC and TGT. In total five out of the six cysteines were coded by a specific codon. For four out of five, TGC was favoured over TGT. This site specific codon bias has only been reported in the variable regions of immunoglobulins (Wagner et al., 1995). It is not known to date how this mechanism functions and which or how many partners are involved.

1.1.3 Conotoxins; receptor and ligand interaction

A number of hypotheses are proposed to explain the interaction of the toxins with receptors such as ion channels (Mok and Han, 1999; Olivera, 1997; Olivera et al., 1991b). One of the hypotheses deals with the question of how different primary sequences can bind with very high affinity to the same binding site. Olivera proposed the *Conus* toxin macrosite model. This is illustrated in Figure 1.1.5. Within the macrosite are potential microsites to which a conotoxin can bind. In a normal situation the endogenous ligand (orange little blobs) can bind to its receptor (orange pocket), which is represented by panel A in the schematic. A conotoxin (blue) blocks the macrosite using three microsites (panel B). Panel C shows a number of possibilities for different conotoxins to bind the same site (blue, red, black).

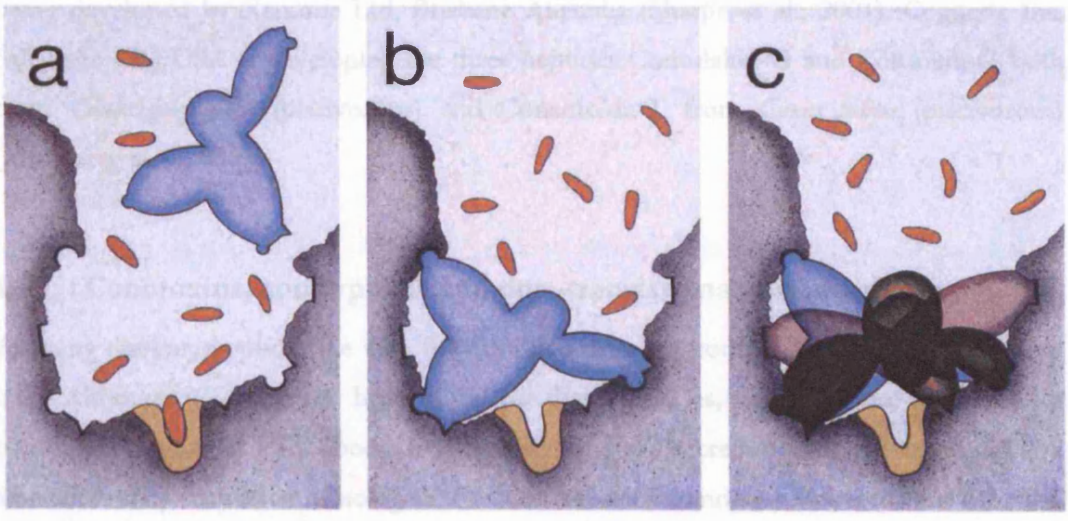


Figure 1.1.5 Conotoxin ion channel binding model.

The schematics in this figure represent the macrosite binding model. Panel A the binding site of a receptor with a number of microsites which are potential inner contacts for conotoxins, the endogenous ligand agonist are the orange blobs. Panel B shows a conotoxin blocking the ligand binding site using three microsites. Panel C shows the same ligand binding site blocked by several conotoxins, making use of the microsites. This figure is reproduced from Olivera, BM *et al.* (Olivera et al., 1991a).

1.1.4 Conotoxins as molecular tools and medicine

A small number of *Conus* venom derived peptides have made it past the laboratory as pharmacological tools and are being developed as pro drugs for the treatment of a range of neurological conditions (Alonso et al., 2003). α-conotoxin Vc.1 from *Conus victoriae* (molluscivorous) is a competitive blocker of selected neuronal-type nicotinic ACh receptors by Metabolic Pharmaceuticals, Melbourne Australia (Sandall et al., 2003). α-conotoxin TIA from *Conus tulipa* (piscivorous) acts as a reversible noncompetitive inhibitor of α-1 adrenergic receptors, developed by Xenome Ltd, Brisbane Australia (Sharpe et al., 2001). ω-conotoxin CVID from *Conus catus* (piscivorous) blocks N-type calcium channel specific sub-type and is being developed by AMRAD, University of Queensland Australia (Blanchfield et al., 2003). ω-conotoxin MVIIA (from *Conus magnus* a piscivorous species) is also known by several commercial names, SNX-111, C1002, Ziconotide and Prialt™. This was the first peptide from *Conus* venom to be utilized as potential pain relief drug. It also blocks N-type calcium channels and it is being developed by Elan Pharmaceuticals, CA USA (Atanassoff et al., 2000; Bowersox and Luther, 1998; Levin et al., 2002; Penn and Paice, 2000). χ-conotoxin MrIA/B from *Conus Marmoreus* (molluscivorous), acts as

reversible noncompetitive inhibitors of the neuronal noradrenaline transporter. These are being developed by Xenome Ltd, Brisbane Australia (Sharpe et al., 2001). Cognetix Inc, Salt Lake City USA is developing the three peptides Contulakin-G and Contokin-G both from *Conus geographus* (piscivorous) and Conantokin-T from *Conus tulipa* (piscivorous) (Malmberg et al., 2003).

1.1.5 Conotoxins, contryphans and post-translational modifications

Amoung the conopeptides (i.e with few disulfide linkages) contryphans are a case on their own. Although no target has been found for these peptides, fish that were injected with contryphans showed elicit body tremor and mucous secretion (Alonso et al., 2003). Therefore they may have a biological function related to endocrine or neuronal function (Massilia et al., 2001).

Contryphans are single disulfide linked members of the conotoxin family (Terlau, 2004) and were first discovered in the venom of the piscivorous snail *Conus radiatus* (Jimenez et al., 1996). Contryphans are highly post-translationally modified small peptides. Typically only 8 to 12 amino acids long, these peptides incorporate modifications such as 4-trans-hydroxy-proline (Stone and Gray, 1982), Leucine (Jacobsen et al., 1999) and Tryptophan (Jacobsen et al., 1998; Jimenez et al., 1996) isomerisation, C-terminus amidation (Nakamura et al., 1996) and bromination of Tryptophan (Craig et al., 1997; Jimenez et al., 1997). The post-translational modification of L- to D-amino acid is catalyzed by a enzyme and has been found in several biological systems (Kreil, 1997).

1.2 Polypeptide Separation

1.2.1 Liquid Chromatography

Many of the experiments carried out for the work in this thesis made use of liquid chromatography. It was used for sample fractionation and as front-end separation before mass spectrometric analysis (both by itself and in combination with gel electrophoresis). Molecules can be separated by chromatography using different properties, such as size, charge and hydrophobicity.

1.2.1.1 Size exclusion chromatography

Size exclusion chromatography (SEC) is a chromatographic method in which molecules are separated based on their size and was first described by Porath *et al.* (Porath and Flodin, 1959). This method is most widely used in the analysis of polymer molecular weights (or molar mass).

Molecules are separated by whether or not they can fit within the pore size of the packing material. When columns are created they are packed with porous beads with a specific pore size so that they are most accurate at separating molecules with sizes similar to the pore size. As a molecule flows through the column it passes by a number of these porous beads. If the molecule can fit inside the pore then it is drawn in by the force of diffusion. There it stays a short while and then moves on. If a molecule can not fit into a pore then it continues following the solvent flow. For this reason, in a SEC column, molecules with larger size will reach the end of the column before molecules with smaller size. The effective range of the column is determined by the pore size of the packing. Any molecules larger than all the pores in a column will be eluted together regardless of their size. Likewise, any molecules that can fit into all the pores in the packing material will elute at the same time.

The only absolute measure in SEC is volume of the molecule (hydrodynamic volume), and even that measurement has a certain error built into it. Interactions between the solvent, packing, and/or the sample will affect the measurement as will concentration due to sample-sample interactions (Kostanski et al., 2004)

1.2.1.2 Reversed phase chromatography

In Reversed phase liquid chromatography (RP-LC), molecules are separated according to their hydrophobicity (Snyder et al., 1997). Sample mixtures in a hydrophilic mobile phase are introduced to a column containing a hydrophobic stationary phase. Elution occurs with applications of a hydrophobic phase applied in a fixed (isocratic) or variable (gradient) concentration.

Depending on the hydrophobicity of the analyte, different stationary phases can be used. More hydrophobic species, such as proteins can be separated by relatively hydrophilic alkylsilane stationary phases such as butyl (C4) or octyl (C8) (i.e. chromatographic silica beads derivatised with aliphatic chains containing 4 or 8 carbon atoms) containing relatively

large pore sizes. In contrast, small peptides are better separated with columns packed with beads derivatised with octadecyl groups (i.e. C18) containing relatively small pore sizes (Snyder et al., 1983)

1.2.2 High performance two-dimensional gel electrophoresis

Two-dimensional polyacrylamide gel electrophoresis (2-DE) is one of the core technologies of proteome analysis (Hoving et al., 2002). Proteins are separated according to charge (pI) by isoelectric focusing (IEF) in the first dimension and according to relative mass (Mr) by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension. 2-DE was first described by O'Farrell (O'Farrell, 1975), Klose (Klose, 1975) and Scheele (Scheele, 1975). It has become the method of choice for the separation of complex mixtures of proteins because of its high resolution, due to the independent parameters for the two dimensions. A major drawback for 2-DE for a long time was reproducibility. This was mainly due to the instability of the pH gradient formed and the limited pH range that could be created (pH 4-8) with carrier ampholytes in the first dimension. Furthermore, preparative amounts of proteins could not be separated very well on these types of gels. The introduction of immobilized pH gradient (IPG) gels largely overcame the problem of gradient stability and poor sample loading capacity (Bjellqvist et al., 1982; Righetti, 1989).

1.2.2.1 The first dimension: isoelectric focusing

The use of isoelectric focusing is limited to molecules that can be positively and negatively charged in a changing pH environment (Westermeier, 1997). Proteins and peptides are such molecules. The net charge of a protein is the sum of all negative and positive charges. The three dimensional configuration of proteins does not play a role as the protein is assumed to be totally denatured (random coil), because of the chaotropic chemicals used in the solubilisation buffer. A post-translational modification, such as phosphorylation has an influence on the net charge of a protein, so this is usually seen at a spot train on the gel. Isoelectric focusing is in principle an end point method. Once the protein has reached the pH where its net charge is zero, it is in equilibrium in a focal zone (Righetti, 1989).

1.2.2.2 Second dimension SDS-PAGE

After the first dimension (IEF), strips are equilibrated for the second dimension. This treatment has three functions. Reduction of disulfide bonds, acetylation and SDS treatment In order to maintain solubilisation in the second dimension, disulphide bonds are once more reduced with dithiothreitol (DTT) according to reaction (1) and alkylated (protein-CH2-S-CH2-CO-NH2) by iodoacetamide according to reaction (2). Alkylation prevents the formation of new disulfide bonds.

$$R\text{-}CH_2\text{-}S\text{-}S\text{-}CH_2\text{-}R + C_4H_{10}O_2S_2 \rightarrow 2\,R\text{-}CH_2\text{-}SH + C_4H_8O_2S_2 \quad (1)$$

$$R\text{-}CH_2\text{-}SH + ICH_2CONH_2 \rightarrow R\text{-}CH_2\text{-}S\text{-}CH_2\text{-}CO\text{-}NH_2 + HI \quad (2)$$

In the second dimension, the proteins are separated on a 20 by 25 cm gel, accorded to their relative molecular weight by SDS-PAGE. The detergent SDS binds to the proteins at a ratio of about one SDS molecule per two amino acid residues in such a way, that they al have the same net negative charge density and migrate in a gel according only to their relative molecular mass. The strips are transferred to second dimension 12 % poly-acrylamide (PA) gels. An electric field is applied and the proteins migrate towards the anode, due to their negative charge. The run is stopped when the tracking dye, bromophenol blue, reaches the end of the gel.

1.2.2.3 Difference Gel Electrophoresis

Difference Gel Electrophoresis (DIGE) is a relatively new 'staining' technique and is used for relative quantitation of 2DE separated proteins. Proteins are labelled on lysine side groups prior to separation with different fluorescent dyes so called CyeDyes (Cy3 and Cy5). A pool of both samples is stained with a third dye (Cy2), this acts as an internal standard. All samples are then combined and run on the same gel. All tree dyes have distinct and non-overlapping excitation and emission spectra enabling their specific detection with a multi-wavelength fluorescent detection device. The dyes should be mass and charge matched and the dye modification should not perturb the electrophoretic mobility of labelled proteins.

1.3 Bioanalytical Mass Spectrometry

The determination of molecular weight is among the first measurements used to characterise biopolymers. Up to the end of the 1970s, the only techniques that provided this information were electrophoresis, chromatographic or ultracentifugation methods (Hoffmann and Stroobant, 2003). The error on the measurement could be 10-100% as all of these methods are not only measuring molecular weight, but the measurement is influenced by parameters like hydrophobicity and protein folding.

Mass spectrometry (MS) has become the technique of choice over the last two and a half decades, for protein identification, and increasingly the investigation and discovery of post-translational modifications (PTM). This is mainly due to a number of instrumental advantages, such as sensitivity, low sample consumption, accuracy, possibility to automate etc (Baldwin, 2004).

1.3.1 Mass Spectrometer, a quick overview

In essence, a mass spectrometer is used to produce ions and measure their mass to charge ratio (m/z) of ions. A mass spectrometer consists of three basic parts, namely (i) an ionization source, (ii) an analyzer and (iii) a detector.

i. Analyte molecules are converted into gaseous ions. The most used ionization sources for biological molecules are Matrix-Assisted Laser Desorption / Ionization (MALDI), where LASER stands for Light Amplification by Stimulated Emission of Radiation. The other most used ionization source is Electrospray Ionization or ESI. Other ionization methods include Fast Atom Bombardment (FAB), Chemical Ionization (CI), thermal Ionization (TIMS), Secondary Ionization (SIMS) and Plasma Desorption (PD).

ii. An electric or magnetic field can deflect charged particles, and since the kinetic energy imparted by motion through an electric field gives the particles an inertia dependent on the particle's mass, the mass analyser uses these facts to steer certain ions to the detector based on their m/z ratios by varying the electrical or magnetic field. It can be used to select a narrow range of m/z (for instance to select peptides of interest for tandem mass spectrometry (MS/MS)) or to scan through a range of m/z to catalog the ions present (survey scan). Examples of mass analysers are

quadrupole mass analyzers, time-of-flight (TOF), ion trap (IT), ion cyclotron resonance (ICR) and magnetic sector instruments. There are numerous combinations of mass analyzers in so called hybrid instruments. The first three are the most commonly used instrument in bioanalytical mass spectrometry, recently also very high accuracy mass spectrometry has become commercially available (Ion Cyclotron Resonance Fourier Transformation Mass Spectrometry or ICR-FT-MS).

iii. There are numerous ways to detect ions. Routinely these are recorded when an ion hits (Multi Channel Plate, or MCP) or passes by a detector plate. As ions hit the plate a cascade of electrons is released, amplifying the single ion detection. As ions pass by the detector plate they cause electrons to flow to the plate. This flow is called image current and can be detected and amplified. If a scan is conducted in the mass analyzer, the charge induced in the detector during the course of the scan will produce a mass spectrum, a record of the m/z values at which ions are present.

1.3.2 Ionization methods

Biological mass spectrometry has and is being developed at a rapid pace since the development of the soft ionization techniques MALDI and ESI. MALDI was developed by Karas and Hillenkamp (Hillenkamp et al., 1991; Karas and Hillenkamp, 1988) and Tanaka (Tanaka et al., 1988). Electrospray was developed by Fenn (Fenn et al., 1989). Fenn received the Nobel Prize for Chemistry in 2002. In the same year Tanaka was awarded the Nobel Prize for his contribution to the development of soft ionization methods for mass spectrometric analysis of biological molecules.

1.3.2.1 Matrix-Assisted Laser Desorption Ionization

Lasers found their first application in mass spectrometry in the early 1970s for desorption / ionizations of small organic molecules that contained appropriate chromophores (Cotter, 1984). The rapid energy absorption initiates a phase change in a microvolume of the absorbing material from solid to gas, while also inducing ionization. In the late 1980s, it was discovered that desorption / ionization of large, non-volatile molecules such as proteins could be similarly effected when such samples were irradiated after being deposited together with a large molar excess of an energy absorbing material (Karas et al., 1987).

The initial experiments were conducted with frequency-quadrupled Neodymium-Yttrium-Aluminium-Garnet (Nd-YAG) lasers that emit at 266 nm, while other investigators have subsequently used longer wavelengths from nitrogen (337 nm) or frequency-tripled Er-YAG lasers (355 nm). Infrared (IR) laser radiation (2.94 μm) from carbon dioxide lasers has also been employed for MALDI-MS. IR lasers are particularly suited for the analysis of labile molecules and protein and peptide modifications (e.g. phosphorylation or glycosylation), because they allow for lower energy transfer to analyte. However, this occurs at the expense of a very high rate of ablation, meaning that a sample is so rapidly desorbed at that frequently that only a single laser pulse can be fired at a particular position before the entire sample is ablated (Cramer and Burlingame, 2000; Cramer et al., 1998).

For the most frequently used laser wavelengths (337 and 355 nm) several hundred compounds have been selected for their radiation absorbing properties for use as MALDI matrices (Ehring et al., 1992; Fitzgerald et al., 1993). Examples of widely used matrices are sinapinic acid, 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydrocycinnamic acid (CHCA). The chemical structure of the latter two is given in Figure 1.3.1.

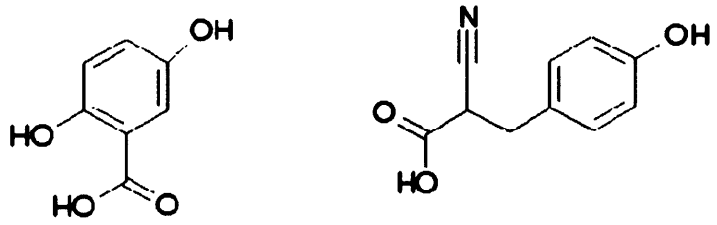


Figure 1.3.1 Examples of commonly used MALDI matrices.

Two commonly used matrices used in MALDI mass spectrometry are 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydrocycinnamic acid (HCCA), show on the left and right respectively. DHB forms a crystal rim, and is relatively salt tolerant. HCCA forms semi homogeneous spot which makes it more amenable for automatic acquisition.

DHB is more tolerant to 'high' salt concentrations compared to HCCA. DHB crystallizes in the form of a spike shaped crystal rim around the centre of the spot. HCCA forms a semi homogeneous shaped matrix spot, making it more suitable for automated data acquisition. Another major difference between these two matrices is the rate of energy transfer to the analyte. DHB is considered a 'cool' matrix as the energy transfer is relatively low, making it more suitable for the analysis of labile modifications on peptides. HCCA on the contrary is a 'hot' matrix. Because of the higher transfer of energy this matrix is more suitable to gain structural information (primary sequence) of peptides. In a recent paper by

Cramer and Corless (Cramer and Corless, 2005) investigate the use of liquid matrices in light of the automatic acquisition of MALDI spectra for high throughput protein determination. This was originally reported by Sze *et al.* (Sze et al., 1998). Liquid matrices generally are conventional matrices dissolved in a viscous, vacuum stable liquid, such as glycerol. The main advantage of this method is that there is no need to sample the entire spot, as the sample remains in solution and is no longer fixed to a specific crystal within the spot.

Figure 1.3.2 shows a schematic of the MALDI process in positive ion mode. The exact mechanism of energy transfer and ionization of analyte molecules is still being debated (Karas and Kruger, 2003; Kocher et al., 2005; Pshenichnyuk and Asfandiarov, 2004; Zenobi and Knochenmuss, 1998). However it is believed that after laser irridiation of the matrix-analyte crystals (h), clusters of matrix-analyte is formed in the so-called plume (d) are formed by a complex combination of thermal desorption, sublimation/evaporation, and ablation by phase explosion induced by the pulsed laser beam (Dreisewerd, 2003). The plume is formed in a so-called field free region before the extraction electrode (E et al.). Only after the formation of the plume, a potential is applied (f) to extract the charged ions from the plume. The delay of the extraction in the process is known as delayed extraction (DE) (Jensen et al., 1996). Not all MALDI instruments are equipped with DE.

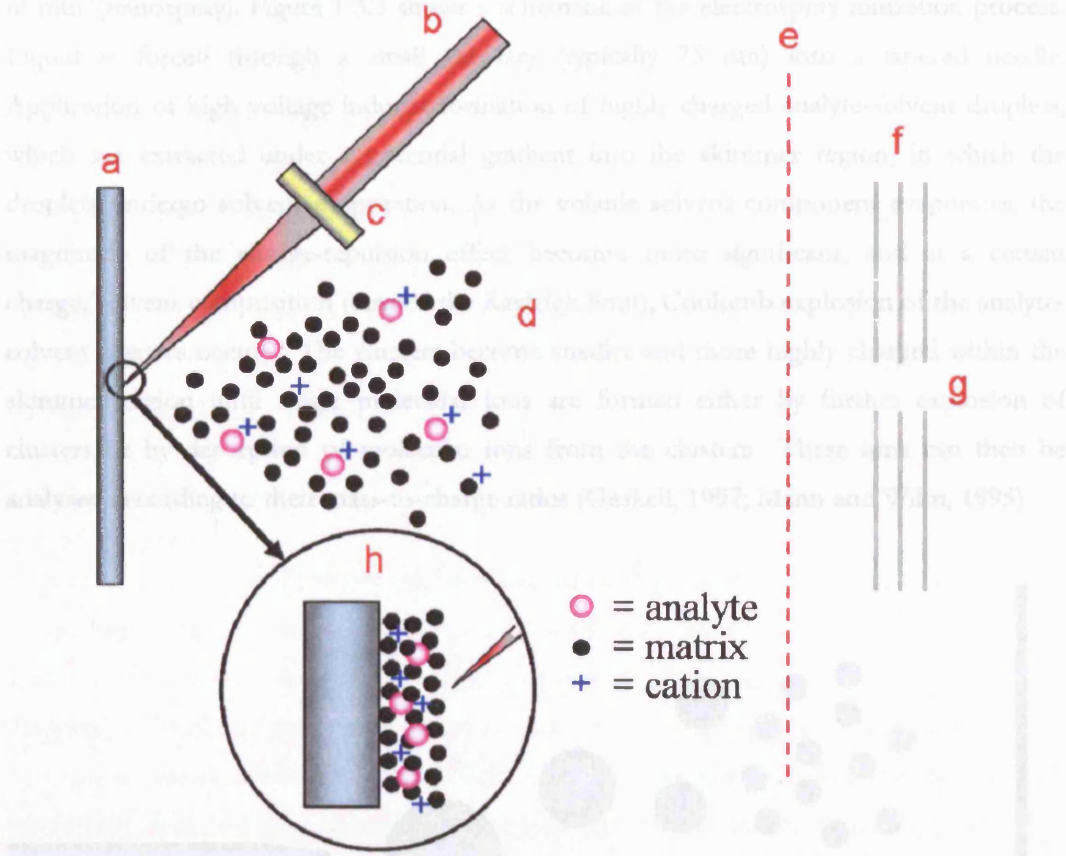


Figure 1.3.2 Schematic representation of the MALDI process

The analyte is co-crystallized with the matrix molecules (h). A pulsed laser (E et al.) is focussed onto the target (a) by ion optics (E et al.). A mixture of matrix and analyte is desorbed into a so called plume (d). The now ionized analyte molecules (magenta) are in the field free drift region before the electrode (E et al.). The charged molecules are extracted by a high potential (f) into the entrance of the mass spectrometer (g).

1.3.2.2 Electrospray Ionization

Electrospray ionisation (ESI) is a very soft ionization method. ESI allows for the formation of ions from liquid phase samples. The sample is dissolved in an organic solvent, typically methanol or acetonitrille, containing a small concentration of acid (formic acid, 0.1 – 1%). The introduction of sample into the mass spectrometer is carried out by a number of methods. In the simplest case, sample is directly infused using a syringe and a transfer capillary. Another example is the so-called nano-spray sample delivery. In this system, a small amount of sample is placed into a needle, which has a very small tapered opening on one side. Sample is forced out because of capillary forces and voltage applied to the conducting needle. However, the most commonly used method is the coupling of electrospray directly with reversed phase chromatography. In this setup the capillary end of the chromatographic system is connected to a needle, with typical flow rates of 200 to 500

nl min^{-1}(nanospray). Figure 1.3.3 shows a schematic of the electrospray ionization process. Liquid is forced through a small capillary (typically 75 μm) into a tapered needle. Application of high voltage induces formation of highly charged analyte-solvent droplets, which are extracted under a potential gradient into the skimmer region, in which the droplets undergo solvent evaporation. As the volatile solvent component evaporates, the magnitude of the charge-repulsion effect becomes more significant, and at a certain charge/solvent composition (termed the Rayleigh limit), Coulomb explosion of the analyte-solvent clusters occurs. The clusters become smaller and more highly charged within the skimmer region until single molecular ions are formed either by further explosion of clusters or by desorption of molecular ions from the clusters These ions can then be analysed according to their mass-to-charge ratios (Gaskell, 1997; Mann and Wilm, 1995) .

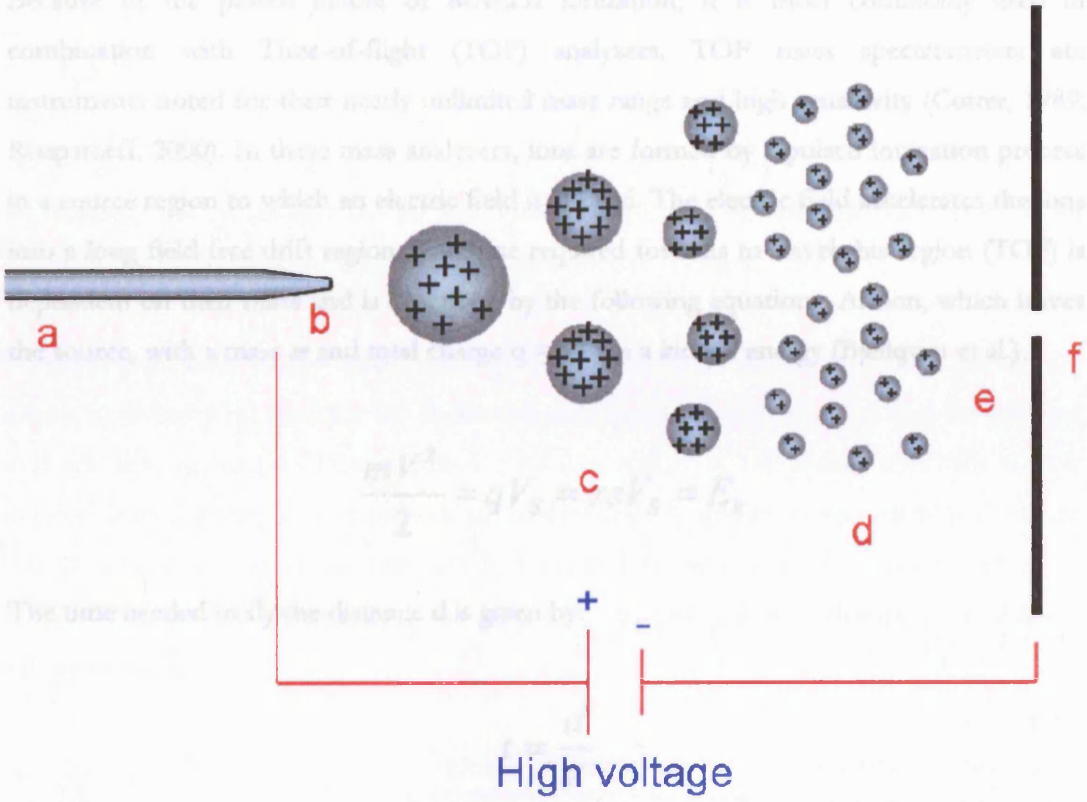


Figure 1.3.3 Schematic representation of electrospray ionization

This figure shows the schematic representation of the ESI process. Sample is delivered through a capillary (a) and a tapered needle (E et al.). Through a high voltage, droplets are extracted (E et al.). Because of the evaporation of solvent the charge repulsion reaches a critical value (the Rayleigh limit) (d), droplets explode (Coulomb explosion), creating generally multiply charged ions (E et al.). These enter the mass spectrometer (f) under the applied electric field.

1.3.3 Mass Analysers and hybrid instruments

Due to the large range of different instruments available, only those mass spectrometers are discussed in detail that were relevant to the work of this thesis.

After producing ions in the ion source, they need to be separated according to their mass to charge ratio. This separation occurs in the mass analyzer part of the mass spectrometer. These mass analyzers can be divided into four general groups (i) time-of-flight (TOF), (ii) quadrupole, (iii) ion trap and (iv) ion cyclotron resonance. As briefly mentioned above, a number of hybrid instruments are commercially available. A few examples of hybrid instrument are discussed below.

1.3.3.1 MALDI-TOF

Because of the pulsed nature of MALDI ionization, it is most commonly used in combination with Time-of-flight (TOF) analyzers. TOF mass spectrometers are instruments noted for their nearly unlimited mass range and high sensitivity (Cotter, 1989; Roepstorff, 2000). In these mass analyzers, ions are formed by a pulsed ionization process in a source region to which an electric field is applied. The electric field accelerates the ions into a long field free drift region. The time required for ions to travel this region (TOF) is dependent on their mass and is described by the following equations. An ion, which leaves the source, with a mass *m* and total charge q = ze has a kinetic energy (Bjellqvist et al.):

$$\frac{mV^2}{2} = qV_S = zeV_S = E_k$$

The time needed to fly the distance d is given by:

$$t = \frac{d}{V}$$

Replacing V (velocity) by its value in the previous equation gives:

$$t^2 = \frac{m}{z}\left(\frac{d^2}{2V_S e}\right)$$

Where V_s is voltage. When the terms in the parentheses are constant, there is a linear relation between the m/z value and time (t^2). Therefore, low-mass ions have a shorter flight time than heavier ions. The start of the measurement for the TOF is triggered by the pulsed laser signal

Because the TOF recorded for a particular ion reflects many different initial conditions experienced in the ion source (such as the time / location of the ion formation and the initial kinetic energy distributions), the mass resolution can be relatively low. Resolution is defined as the reciprocal of resolving power, which is defined as the smallest mass difference between equal magnitude peaks such that the valley between them is a specified fraction of either peak height, or

$$R = \frac{m}{\delta m}$$

To combat the phenomenon of low mass resolution, modern MALDI-TOF mass spectrometers are equipped with a number of hardware features.

To lower the kinetic energy spread among ions with the same m/z ratio leaving the source, a time lag or delay between the ion formation and extraction can be introduced. In this case, ions are first allowed to expand into a field-free region in the source and after a user defined delay (varying from hundreds of nanoseconds to several microseconds) a voltage pulse is applied to extract the ions outside the source (Jensen et al., 1996; Kaufmann et al., 1996). This process / technique is known as delayed pulsed extraction, or delayed extraction (DE).

Another hardware tool, used to improve mass resolution is the electrostatic reflector, or reflectron for short. The reflectron consists usually of a series of grids and ring electrodes (Cornish and Cotter, 1993). It creates a delaying field that acts as an ion mirror by deflecting the ions and sending them back thought the flight tube. The reflector corrects the energy dispersion of the ions leaving the source with the same m/z ratio. Ions with more kinetic energy will penetrate the reflectron more deeply. Therefore they reach the detector at the same time as slower ions of the same m/z. As discussed above, the major

advantage is the enhanced mass resolution, but there is a trade-off in terms of sensitivity and a limited mass range (Hoffmann and Stroobant, 2003).

In recent years tandem TOF mass spectrometers have been developed. These MALDI based instruments are capable of fragmenting ions of choice in a data dependent manner. Figure 1.3.4 shows a schematic overview of a, so called TOF-TOF mass spectrometer. In MS mode, ions are measured as in a conventional MS. In MS/MS, mode an ion of interest is filtered through by the timed ion selector (E et al.), before being collided with a gas in the collision cell (d). Fragment ions are re-accelerated by the second ion source (E et al.) and analyzed by the second TOF, the MS/MS is based on collision induced dissociation (CID).

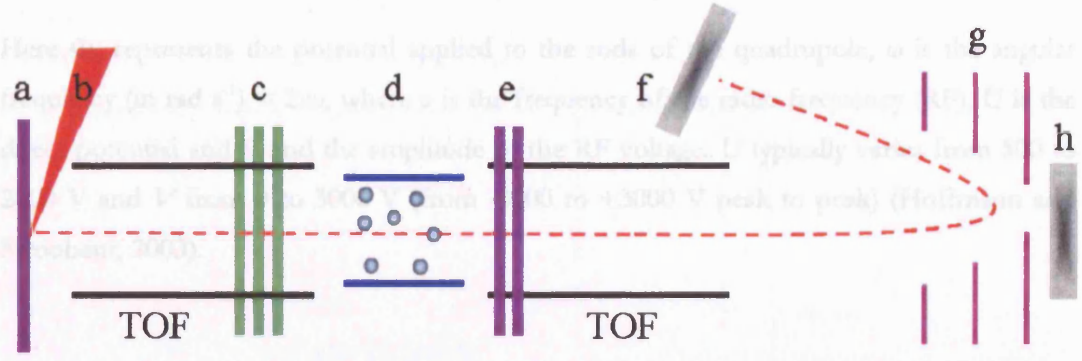


Figure 1.3.4 Schematic overview of a MALDI TOF/TOF instrument

This figure shows the schematic overview of a TOF/TOF mass spectrometer. Ions of interest are gated through the timed ion selector and fragmented by collision induced dissociation before being analysed by the second TOF. The MS consists of the following components: (a) ion source 1 (E et al.) laser beam (E et al.) timed-ion-selector (d) collision cell (E et al.) ion source 2 (f) reflector detector (g) reflector (h) linear detector. The ion path is indicated by the red dashed line.

TOF analyzers are directly compatible with pulsed ionization sources. However, on modern hybrid mass spectrometers, TOF can be used with continuous ion sources, such as electrospray. In that case, packets of ions are retarded by a decelerating field. They are then extracted into the flight region which is perpendicular to the original ions path (Boyle and Whitehouse, 1992).

1.3.3.2 Quadrupolar analyzers

The quadrupole principle was first described by Paul *et al.* (Paul and Steinwedel, 1953). The quadrupole, the actual mass analyzer, makes use of the stability of the trajectories in

oscillating and static electric fields to separate ions according to their m/z ratio. Quadrupole analyzers are made up out of four perfect parallel inert metal rods. When, for instance, a positive ion enters the zone between the rods, it will be drawn towards the negative pole. When the potential is altered before the ion collides with the rod, the ion will change direction.

Ions travelling along the z-axis are subjected to the influence of a total electric field made up of a quadrupolar alternating field superposed on a constant field resulting from the application of the potentials upon the rods:

$$\Phi_0 = +(U - V\cos\omega t) \quad \text{and} \quad -\Phi_0 = -(U - V\cos\omega t)$$

Here Φ_0 represents the potential applied to the rods of the quadrupole, ω is the angular frequency (in rad s^{-1}) = $2\pi\upsilon$, where υ is the frequency of the radio frequency (RF), U is the direct potential and V and the amplitude of the RF voltage. *U* typically varies from 500 to 2000 V and *V* from 0 to 3000 V (from -3000 to +3000 V peak to peak) (Hoffmann and Stroobant, 2003).

1.3.3.3 Orthogonal Quadrupole TOF

Multiple quadrupoles can be used in a mass spectrometer. Q-TOF type instruments are an example of such an instrument, combined with time of flight mass analyzer. The Q-TOF was developed by Morris *et al.* (Morris et al., 1997). The first commercial instrument Q-TOF™ was developed by Micromass. These types of instruments have three tandem quadrupoles, of which the second (q2) is an RF-only quadrupole. In RF-only quadrupoles there is no direct potential, therefore ions of certain m/z ratio, above a lower limit have a stable trajectory, as long as V is within the limit of their stability area. Such a quadrupole causes all of the ions to be brought back systematically to the centre of the rods, even if they were deflected by a collision. This is called ion focusing and is important to increase the transmission of ions through the mass spectrometer after collisions. Figure 1.3.5 shows a schematic of a Q-TOF type instrument. As discussed above, in case of continuous electrospray, there is a need to place the TOF orthogonal to the ion path. Packets of ions are 'pushed' into the flight tube, initialising the start of a time of flight measurement. Many manufacturers have also developed a MALDI ionization source for these instruments.

In general, quadrupoles are indicated by 'Q', while RF-only quadrupole are indicated with 'q'. This hybrid instrument consisting on a quadrupole (Q1) and a TOF connected through another quadruple (q2) that is used as a collision cell. When scanning a large mass range, Q1 and Q2 operate in RF mode only so that ions with a wide m/z range are transmitted to a push-pull region were ions are accelerated orthogonally towards the field-free region of the TOF. In MS/MS experiments Q1 isolates the desired ion and transmits it to q2 where it fragments due to collision induced dissociation (CID). Ions of interest are selected in a data dependent manner. For this, parent masses are automatically selected for tandem mass spectrometry (MS/MS) when, for instance, the signal in the MS spectrum rises above a user defined threshold. Q3 focusses the ions towards the push-puller region (a).

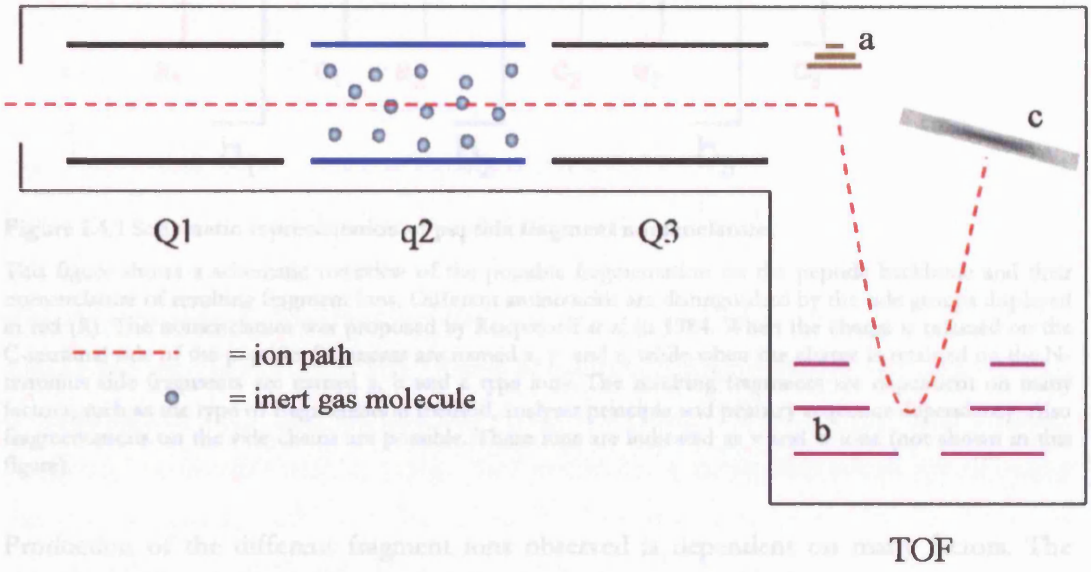


Figure 1.3.5 Schematic overview of a qQ-TOF mass spectrometer

This figure gives a simplistic overview of an orthogonal Q-TOF mass spectrometer. Quadrupoles are generally indicated by letter 'Q' and RF-only quadrupoles with 'q'. In this experiment q2 is filled with an inert gas and functions therefore as a collision cell. a) is a push region that pushes the ions into the orthogonal placed TOF. b) reflector and c) detector plate. The ion path is displayed by the dashed red line.

1.4 Peptide Sequencing using Tandem Mass Spectrometry

The main objective of tandem mass spectrometry (MS/MS) is to gain structural information. Figure 1.4.1 shows a schematic representation of the possible fragmentation of the peptide backbone. The nomenclature was proposed by Roepsdorff *et al.* (Roepstorff and Fohlman, 1984). Upon fragmentation of a peptide the charge is retained either by the N- or C- terminus side of the molecule. Fragments are named a, b or c in the first case or x,

y or z in the latter. Fragmentation induced by high collision energy result in, so called v and w ions, that occur on the side chain groups and are not discussed here.

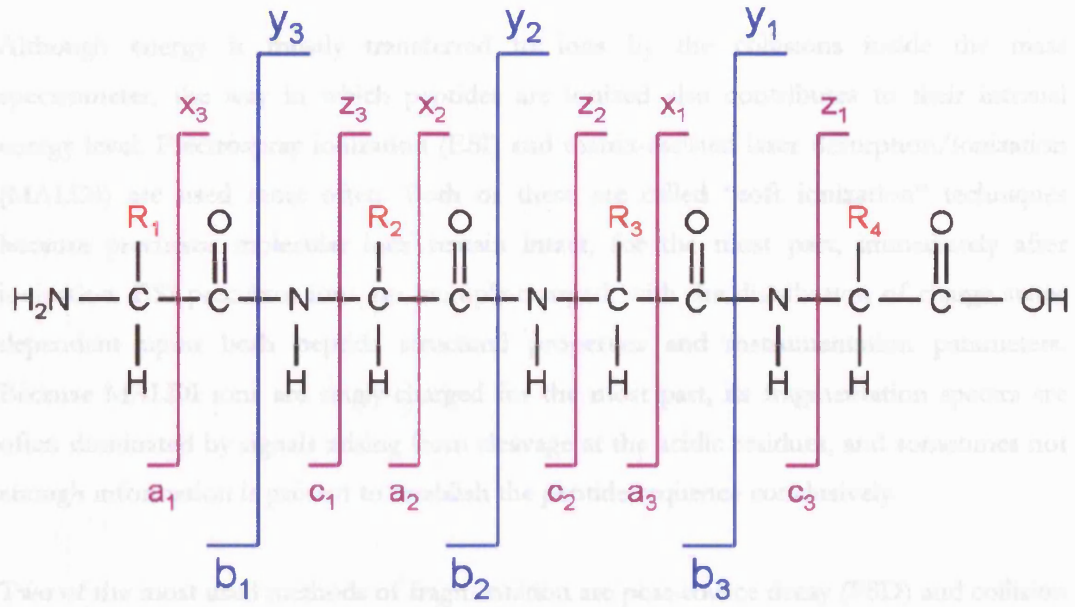


Figure 1.4.1 Schematic representation of peptide fragment nomenclature.

This figure shows a schematic overview of the possible fragmentation on the peptide backbone and their nomenclature of resulting fragment ions. Different amino acids are distinguished by the side groups displayed in red (R). The nomenclature was proposed by Roepstorff *et al.* in 1984. When the charge is retained on the C-terminal side of the peptide, fragments are named x, y and z, while when the charge is retained on the N-terminus side fragments are named a, b and c type ions. The resulting fragments are dependent on many factors, such as the type of fragmentation method, analyser principle and primary sequence dependency. Also fragmentations on the side chains are possible. These ions are indicated as v and w ions (not shown in this figure).

Production of the different fragment ions observed is dependent on many factors. The mobile proton model is a framework for the understanding of peptide dissociation and was described by Wysocki *et al.* (Wysocki et al., 2000). According to this model, fragmentation of most protonated peptides requires a proton to be proximal to the cleavage site. Peptides usually fragment at their amide bond (peptide bond) producing b and y ions. The mobile proton model also states that the cleavage is initiated by the migration of the charge from the initial site of protonation. If the side chain on a basic amino acid tightly binds the proton, energy must first be delivered to the ion to shift the proton to the peptide backbone, which then induces dissociation. The exact amount of energy required depends on the amino acid composition. Arginine containing peptides need more than lysine-containing peptides, which, in turn, need more energy than peptides with no basic residues. Magnetic sector instruments and tandem time-of-flights (TOF/TOFs) can induce higher energy collisions for investigation of these alternative pathways. Not surprisingly, the

peptide sequence is the primary determinant of how much energy is required for optimal fragmentation.

Although energy is mostly transferred to ions by the collisions inside the mass spectrometer, the way in which peptides are ionized also contributes to their internal energy level. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are used most often. Both of these are called "soft ionization" techniques because precursor molecular ions remain intact, for the most part, immediately after ionization. ESI precursor ions are multiply-charged, with the distribution of charge states dependent upon both peptide structural properties and instrumentation parameters. Because MALDI ions are singly-charged for the most part, its fragmentation spectra are often dominated by signals arising from cleavage at the acidic residues, and sometimes not enough information is present to establish the peptide sequence conclusively.

Two of the most used methods of fragmentation are post-source decay (PSD) and collision induced dissociation (CID). PSD is a common side effect, often seen in MALDI mass spectrometry. As the name indicates, this sort of fragmentation (metastable decay) occurs after the formation of the actual ion. PSD was first utilised by Kaufmann *et al.* (Kaufmann, 1995; Kaufmann et al., 1993). The process of metastable fragmentation following MALDI is employed in PSD analysis, generally using a reflectron time-of-flight mass analyser. Fragmentation is maximised by using a 'hot' matrix (i.e. a matrix with a high rate of energy transfer to analyte, such as HCCA), and by increased laser power, which promotes fragmentation. Fragment ions are resolved by altering the focussing of the reflectron over small regions of the m/z range to maximise the resolution and mass accuracy of the fragments. The PSD fragmentation process produces many internal peptide fragment ions (bearing neither N- or C-terminus of the peptide), as well as peptide backbone cleavages. Such types of ions typically make interpretation of sequence less facile than that of CID data. The LIFT-TOF instrument from Bruker Daltonics makes use of this principle for tandem MS. A peptide of interest and its PSD fragments (both have the same initial velocity, independent of mass) are gated through by the timed ion selector. After the first TOF tube ions are re-accelerated into the second TOF.

CID is a more widely used fragmentation technique, as is it used both on MALDI and ESI ionization mass spectrometers. Peptides of interest are selected for tandem MS and

accelerated into a collision cell. This is usually filled with an inert gas. After collision ions are re-accelerated.

In both cases, a-, b- and y- ions are the predominant fragment ions. Tandem MS data is more and more used as a standard to interrogate databases. With the increasing amount of genetic data available for different organisms, databases steadily grow in size. In the beginning of the 90s it was often enough to use peptide mass fingerprinting (PMF), this is a technique whereby measured masses of peptides generated by proteolytic cleavages are searched against a theoretical digest from proteins in a database. Nowadays more accurate data is necessary to assign proteins IDs. MS/MS data can also be used to assign novel, unknown peptides. This process is referred to as de novo sequencing.

When performing de novo sequencing, many other diagnostic signals from the mass spectrum are used. Two of these diagnostic fragment ions are shown in Figure 1.4.2. Amino acids like proline and carbamidomethylated cysteine are prone to generate internal fragment ions. These are the result of a double cleavage of the peptide backbone and the central part retaining the charge. When an internal fragmentation occurs on the y- and a-ion of the same amino acid, an immonium ion is generated. A schematic representation is shown on the right side of the Figure 1.4.2.

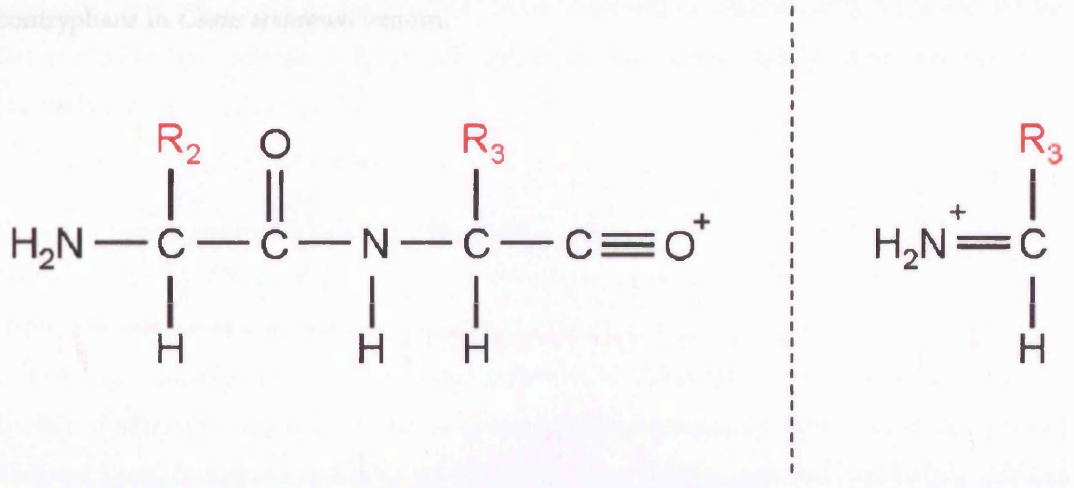


Figure 1.4.2 Schematic representation of internal fragments and immonium ions.

This figure shows a schematic representation of an internal fragment ion (left) and an immonium ion (right). Internal fragment ions occur when the peptide backbone is cleaved in two places and the charge is retained by the central part of the peptide/ion. Amino acids like proline and carbamidomethylated cyteine are prone to generate such fragments. An immonium ion is a diagnostic ion specific to an amino acid. It is the result of a y- and a ion cleavage.

1.5 Aims and scope

The primary aim of this study was to set up a separation and analysis scheme to rigorously investigate the components of *Conus* venom with Mass Spetrometry as a central technique. As discussed above, *Conus* derived toxins are very potent drugs that act mainly on ion channels and other cell receptors. Although conotoxins have been investigated for more then two decades, only a handful of the estimated 50,000 potential biologically active compounds have been characterised at the protein/peptide level. This work is not concerned with biological assays of the components of *Conus* venom.

In this work several strategies have been employed to investigate the complexity of *Conus* venoms. Chapter three is concerned with the analysis of the protein content of venom samples. Both one and two-dimensional gel electrophoresis was used.

Chapters four, five and six describe the analysis of the peptide content of the different venoms available. A size exclusion method, followed by LC-MALDI-MS/MS was performed on the venom of *Conus ventricosus* and is described in Chapter four. This method revealed several partial conotoxin sequences. Chapter five describes the separation and analysis of *Conus ventricosus* and *Conus textile* venom, where size separated fractions were further analysed by LC-ESI-MS/MS. Chapter six shows the discovery of several new contryphans in *Conus ventricosus* venom.

2 Materials and Methods

2.1 Mass spectrometry

Mass spectrometry based experiments were carried out using instruments located and maintained at the Bioanalytical Chemistry laboratory (Professor Dr R. Cramer, group leader) of the Ludwig Institute for Cancer Research (UCL branch, Prof M.D. Waterfield, director) or in the Mass Spectrometry Facility of the University of California in San Francisco (Professor Dr. A.L. Burlingame, director).

2.1.1 MALDI-TOF MS and MALDI-TOF/TOF MS/MS

Although a suite of MALDI mass spectrometers is available in both labs, all MALDI data presented in this thesis was obtained on a 4700 Proteomics Analyzer MALDI-TOF/TOF (Applied Biosystems, Foster City, CA, USA). This is a low pressure MALDI-TOF/TOF capable of high collision energy fragmentation in a collision cell located in between two TOF analysers.

In MS mode, 50 sub-spectra of 50 laser shots each were accumulated, for MS/MS 400 sub-spectra of 50 laser shots each were collected. Automatic MS/MS was stopped before reaching the 400 sub-spectra if at least 30 sub-spectra contained 18 or more peaks with signal to noise rations above 24. For MS/MS a transmitting window of ± 3 Da was set for the precursor ion selection. Each MS spectrum was automatically calibrated using an internal standard (ACTH 18-39).

2.1.1.1 Sample preparation for MALDI analysis

Samples for MALDI-TOF MS analysis were prepared using the dried droplet protocol. In short, the sample analyte was mixed with an excess of matrix. Typically, 0.5 µl of sample solution was mixed with 1 µl of saturated solutions of 2,5-dihydroxybenzoic acid (2,5-DHB, Bruker Daltonics or Sigma) or α-cyano-4-hydroxycinammic acid (CHCA, Agilent Technologies). Saturated solutions of 2,5 DHB were freshly prepared just before analysis by dissolving an excess of solid matrix in HPLC grade water. HCCA was purchased as a ready made solution in methanol. Samples in DHB were dried under a stream of cold air. HCCA matrix used for automatic spotting with the Probot (LC Packings, San Francisco)

consists of 1 part 70 % methanol and 1 part ready HCCA solution (Agilent Technologies) ACTH 18-39 is added to a final amount equal to 4 fmol per spot.

2.1.2 ESI - QTOF MS

Electrospray MS data for two-dimensional liquid chromatography and 1D gel separation of proteins from *Conus* venom were generated on a QSTAR™ (Sciex / Applied Biosystems, Foster City, CA, USA). Measurements of in-gel digests from 2-DE were performed on a Q-Tof™ instrument (Micromass, Manchester, UK).

2.1.2.1 Operation of ESI-Q-Tof

The QSTAR™ was equipped with a nanoflow ion source. Samples are introduced either by direct infusion by online coupled HPLC through a 20 µm capillary. The Q-Tof™ was equipped with nanoflow ion sources and a fused silica ESI emitter with 50 µm I.D. tapered to 15 µm at the tip (PicoTipTM, New Objective, Woburn, MA). In both methods, the tip and the capillary are not coated with conductive material and the high voltage was applied at a liquid metal junction about 3 to 5 cm from the tip of the emitter. Flow rates for both methods are typically between 200 and 350 nl min^{-1}.

Optimal operation parameters were determined empirically by infusing a standard peptide. Typical spray tip potentials were 1800-3500 V depending on the tip of the emitter and the flow rates. Voltages at the orifice of the interface ranged from 40 to 150 V. The QSTAR™ and Q-Tof™ instruments were calibrated using the m/z values of the fragment ions generated from a MS/MS fragmentation of [Glu1]-Fibrinopeptide B.

2.1.3 LC-MALDI-MS

Before MALDI-MS/MS the protein digests were separated on an Ultimate™ HPLC system (LC Packings, Amsterdam, Netherlands). This consists of an autosampler (Famos™) that loads 5 µl of sample solution at a flow rate of 40µl/min onto a peptide trap column (300 mm inner diameter, 5 mm length, 100 Å particle size PepMap, LC Packings) connected to at a switching valve (Switchos™). Peptides loaded in the peptide trap column were washed for 2 minutes at 40 µl/min with 0.1% formic acid. After this time, peptides were eluted using gradient elution at 200 µl/min onto a main analytical column (C18 PepMap, 75 µm x 15 cm).

Solvent A was 3% Acetonitrile (ACN) 0.1% formic acid and solvent B was 80% (ACN) / 0.1% formic acid. Elution was carried out from 5%B to 45%B in around 60 minutes, after which the column was flushed with 100% B. Before starting a new elution, the column was equilibrated for 20 minutes with 100% A.

The eluent was spotted onto a stainless steel MALDI plate which was placed on a x-y stage robot (Probot, LC Packings\Dionex, San Francisco). The capillary end of the chromatographic system was embedded into a hollow needle. 0.5 µl HCCA matrix (as described above) were dispensed every 30 second, in which the eluent flows (30 seconds sampling rate). The matrix solution is spiked with an internal standard (IS), namely a peptide fragment of adrenocorticotropic hormone (ATCH 18-39). The samples were allowed to dry by air before being analysed. Plates were analysed using the 4700 Proteomics Analyzer MALDI-TOF/TOF (PE Biosystems, Foster City, CA, USA) as described above.

The IS was used to calibrate each TOF-MS spectrum automatically. 5,000 shots were used to generate each MS spectrum. The 8 most intense peaks of each TOF-MS spectrum were automatically selected for MS/MS, providing that the signal to noise ratio was over 25. Parent ions were fragmented for 10,000 shots. The MS/MS was stopped when 18 peaks reached a signal to noise ratio above 25.

2.1.4 LC-ESI-QTOF-MS

Nanoflow LC was performed on an Eksigent nanoHPLC pump (Eksigent, Livermore, CA) linked to a FAMOS™ autosampler (LC Packings/Dionex, San Francisco, CA) or an Ultimate™ HPLC system (LC Packings, Amsterdam, Netherlands).

The Eksigent system, unlike other nanoflow HPLC systems does not use a flow split. The flow is controlled by pressure rather then a pump. The Famos™ autosampler delivered typically 1 µl sample into the sample loop. This is directly loaded onto the main analytical column (C18 PepMap, 75 µm x 15 cm). The Ultimate™ HPLC system were as described above. With both systems a similar gradient was used as decribed above.

The eluent of the HPLC runs were analysed on-line by one of the QTOF mass spectrometers described above. When the QSTAR™ or the QTof™ was used they were operated in data dependent acquisition, which allows for the automatic switching from MS

to MS/MS experiments whenever an ion of a predetermined nature is detected. Briefly, a MS-survey scan is recorded. When an ion reaches a certain ion count a tandem MS experiment is performed, after which the ion (m/z value) is placed in an exclusion list to prevent multiple MS/MS on the same ion. This process is then repeated until the end of the chromatographic elution.

2.2 Liquid chromatography

2.2.1 Size exclusion chromatography

Size exclusion experiments were carried out on an AKTA HPLC (Amersham Biosciences) or a GoldStar (Beckman Coultier). The samples used for the experiment described in Chapter 4 were separated on a TSKfel Super SW2000 column (Tosoh Bioscience). Proteins were eluted in an isocratic run, with a flow rate of 0.5 ml min^{-1} using 25 mM Ammonium Bicarbonate (ABC). Fractions were collected manually in between two absorbance peaks. For all other size exclusion separations, a Tricorn Superdex Peptide 10/300 GL (Amersham Biosciences) column was used. The flow rate was set at 0.5 ml min^{-1} using 150 mM NaCl / 20 mM Tris (pH 7.2) as the eluting buffer. No gradient was applied. Fractions were collected automatically every second minute.

2.2.2 Column packing

For the MALDI-MS experiments, columns were used made in-house. Before packing glas capillaries were fritted using potassium silicate solutions that formed a porous plug at the end of the capillary (performed by Dr. Kirk Hansen, UCSF). For this, 300 µl potassium silicate solution (SiO_2:K_2O, 21:9, Merk) was mixed with 100 µl formamide, vortexed for 1-2 minutes, centrifuged, and the solution positioned inside the capillary by capillary action; the end of the capillary was inserted in the solution for 3-4 s. Polymerisation was performed by placing the capillary in a heat block at 37°C for 1 h.

A slurry of the reversed phase material was made by mixing a spatula tip of the material with an organic solvent, typically methanol. This slurry was placed at the bottom of a in-house constructed pressure cell. The non-fritted end of the capillary was placed at the end of a slurry reservoir through the lid of the pressure cell using standard valco connections. The whole cell was placed user pressure with nitrogen at approximately 800 psi. While

packing the column, the slurry was mixed with a micro magnetic stirrer and the capillary was agitated from time to time.

2.2.3 Reversed phase HPLC

2.2.3.1 Reversed phase clean up of size exclusion fractions

Fractions from size exclusion separation (peptide fractions) were concentrated under vacuum from 1 mL to approximately 100 µl. These were loaded onto a peptide "captrap" (Michrom BioResources at a flow rate of 0.2 ml min^{-1} using 3 % ACN / 0.2 % TFA. The sample was flushed for 10 to 15 minutes, depending on the UV absorbance (214 nm) reaching previous baseline level. Peptides were eluted stepwise with 50 % ACN and 90 % ACN. This process was repeated twice.

2.2.3.2 Second dimension reversed phase fractionation of Conus *venom*

A Vydac C18 reversed phase column (ID 4.6 mm; length 250 mm) was used for second dimension separation. At a flow rate of 0.5 ml min^{-1} peptides were eluted using a standard H_2O/ACN buffer system. Briefly, the main eluting window is from 10 to 80 minutes were the concentration of ACN is increased from 3 to 50 %. The steepness of the gradient was increased by going to 90 % ACN in 10 minutes. Trifluoro acetic acid (TFA) was used as the ion pairing agent. The concentration was kept constant throughout the elution at 0.1 %. Fractions were collected automatically every 0.6 ml. This fractionation was performed on the AKTA HPLC system (Amersham Biosciences).

2.3 Gel electrophoresis

One-dimensional gel electrophoretic (1DE) separation of *Conus* venom proteins was carried out at the mass spectrometry facility in UCSF San Francisco, while the two-dimensional separation of *Conus* venom was carried out at the Bioanalytical chemistry laboratory LICR, UCL. These are briefly outlined below.

2.3.1 One-dimensional SDS-PAGE

1D-SDS-PAGE was performed with commercially available precast 10 – 20 % gradient Tris-glycine gels, (BioRad). Protein samples were mixed with sample buffer (Table 2.3.1) and boiled for 5 minutes at 100 °C. Gels were run at a constant current of 50 mA, until the Bromo Fenol Blue front reached the bottom of the gel, typically 1 to 2 hours.

Table 2.3.1 Concentration of components in 1DE sample buffer

Reagent	Final concentration
Tris pH 6.8	134 mM
SDS	4% (w/v)
Bromophenol blue (BPB)	0.06% (w/v)
Dithiothreitol (DTT)	6% (w/v)
Glycerol	20% (w/v)

2.3.2 Two-dimensional PAGE

Samples for 2DE were dissolved in 8M urea / 2M thiourea / 4% CHAPS / 10 mM Tris pH 7.2, mixed with ampholytes and DTT (85 mM final concentration), and applied to immobilised pH gradient (IPG) gel strips pH 3-10 (Amersham Biosciences, UK). The strips were allowed to rehydrate overnight at room temperature.

For two 2DE, 24 cm long non-linear pH 4-7 immobilised pH gradient (IPG) strips were rehydrated with labelled samples. Proteins were separated in the first dimension according to their isoelectric point (pI). Iso-electrofocusing was performed in a Multiphor II apparatus (Amersham Biosciences, UK) using the program in Table 2.3.2

Table 2.3.2 Program used for the isoelectric focussing of proteins in IPG strips

Step	Time (hours)	Conditions
1	0.01	300 V, 5 mA, 10 W
2	0.5	300 V, 5 mA, 10 W
3	3	3500 V, 5 mA, 10 W
4	21	3500 V, 5 mA, 10 W
5	25	500 V, 5 mA, 10 W

After focusing, strips were equilibrated in 6M urea / 0.1 M Tris pH 6.8 / 30% glycerol / 1% SDS containing 5 mg/ml (65 mM) dithiothreitol DTT for 10 minutes followed by incubation in 45 mg/ml (240 mM) iodoacetamide (IAA) in the same solutions.

The second dimension was carried out using an Akta™ gel electrophoresis system (Amersham Biosciences) that allowed for running six gels in parallel with dimensions 24 cm x 25 cm x 1.5 mm. Strips were transferred onto 12% polyacrylamide gels cast between low-fluorescence glass plates. Reference markers were placed on one plate prior to bonding

and casting. Gels were bonded to the inner surface of one plate (with reference markers) at casting by pre-treatment with bind-silane solution. The second dimension was run under standard SDS gel based conditions, namely 10 mA per gel for the first 30 minutes and approximately 50 mA per gel until the bromo phenol blue front run of the edge of the gel.

2D gels were scanned between low-fluorescence glass plates. The photomultiplier tube voltage was adjusted on each channel (Cy2, Cy3 and Cy5) for preliminary low-resolution scans to give maximum pixel values within 5 - 10% for each Cy-image but below the saturation level. These settings were then used for high-resolution (100 μm) scans of all gels. Images were exported as tiff files for image analysis.

Images were curated and analysed using DeCyder™ software. For DeCyder™ analysis, the difference in-gel analysis (DIA) module was first used to define spot boundaries and to measure normalised spot volumes of Cy2- ,Cy3- and Cy5-labelled features for calculation of difference ratios for samples run on the same gel. Features resulting from non-protein sources (e.g. dust particles and scratches) were filtered out. The biological variation analysis (BVA) mode of DeCyder™ was then used to match all pairwise image comparisons from DIA for comparative cross-gel statistical analysis. User intervention was required at this stage to set landmarks on gels for accurate cross-gel matching. Comparison of test spot volumes with the corresponding standard spot volumes gave a standardised abundance for each matched spot and values were averaged across triplicates for each experimental condition. Strict criteria for selection of spots displaying altered abundance were applied. Essentially, only spots displaying a $\geq$ 5 average-fold increase or decrease in abundance, matching across all gel images and having a p value <0.05, were selected for identification.

2.3.3 CyDye™ labeling of *Conus* venoms

A sample amount of 150 μg (lyophilised venom) was aliquoted into tubes for CyDye™ labelling. Equal amounts of protein from each sample were mixed to create an internal standard which was labelled with Cy2. Enough material was provided to run the internal standard on every gel. Samples were labelled by addition of 4 pmol of the appropriate CyDye™ per μg of protein (600 pmol/150 μg). Samples were incubated on ice in the dark for 30 minutes. Thorough mixing at all steps was conducted to avoid non-uniform labelling. Reactions were quenched by adding a 20-fold molar excess of L-lysine with incubation on ice in the dark for 10 min. A suitable amount (150 μg) of the Cy2-labelled pool was added

to each gel. Samples were reduced by DTT and carrier Ampholines/Pharmalyte (2%) and some bromophenol blue was added. The volume was adjusted to 450 µL with lysis buffer, the samples agitated for 2 min at room temperature prior to clean up of the samples by centrifugation at 13000 RPM for 10 mins.

2.3.4 Gel staining and analysis

Proteins in 1DE gels were visualised by colloidal coomassie blue using coomassie G-250 dye as described in Neuhoff et al. [165;166]. Briefly, gels were fixed from 3 h to overnight in 50% (v/v) ethanol / 2% (v/v) phosphoric acid. After washing the gels three times with water (half hours per wash), gels were incubated with staining solution (34% (v/v) MeOH / 17% (w/v)$(NH_4)_2SO_4$ / 3% (v/v) phosphoric acid) for 1hour, after which time 0.7 g/L (solid) Coomassie Blue G-250 was added to the solution. Gels were then left shaking until bands of the required intensity were visible. The end point of labeling is reached after 3-4 days of incubation. Sensitivity was thought to be approximately 10-50 ng of protein (BSA).

2.4 Solubilisation of Conus venoms

Venoms from *Conus ventricosus* and *Conus textile* were solubilised using artificial sea water (ASW). ASW is made up according to Table 2.4.1 (Yoon et al., 2004)

6.3 µg lyophilised *Conus ventricosus* venom was mixed with 175 µL ASW. It was left shaking for 1.5 hours (750 rpm) at 4 °C. The sample was spun down and the supernatant was removed. The supernatant had an orange/yellow colour. Another 175 µL ASW was added to the pellet and was left shaking for 1 hour (750 rpm) at 4 °C. This process was repeated one more time. The supernatants were combined to give approximately 520 µL solubilised venom. The same procedure was repeated with 4.6 µg lyophilised *Conus textile* venom.

Table 2.4.1 Components and concentrations of artificial sea water.

Compound	Concentration (mM L^{-1})
Sodium Chloride (NaCl)	460
Potassium Chloride (KCl)	10
Calcium Chloride ($CaCl_2$)	11
Magnesium Chloride ($Mg\ Cl_2$)	55
Tris pH 7.2	20

2.5 Enzymatic digestion of proteins and peptides with trypsin

Trypsin specifically hydrolyzes peptide bonds at the carboxylic sides of lysine and arginine residues. Trypsin is classified as a serine protease. Unmodified trypsin is subject to autolysis, generating fragments that can interfere with protein sequencing, HPLC or mass spectrometry analysis of the peptides. In addition, autolysis can result in the generation of pseudotrypsin, which has been shown to exhibit an additional chymotrypsin-like specificity (Keil-Dlouha et al., 1971). Promega's Trypsin has been modified by reductive methylation, rendering it extremely resistant to autolytic digestion (Rice et al., 1977). In functional stability tests, modified trypsin retains at least twice as much activity as unmodified trypsin after 3 hour incubation at 37°C[3].

To prevent contamination that interferes with mass spectrometric analysis, mainly keratin, gloves and sleeve protectors were worn at all time. The in-gel digestion of 1D-PAGE was performed in a laminar flow hood. Surfaces in the hood, including the outside of tubes and pipettes were cleaned with 70% methanol.

For 2D-IEF-PAGE experiments, sample solubilisation, sample labelling, running of first and second dimension gels, staining, spot excision and digestion were performed in a dedicated clean room.

2.5.1 In-gel trypsin digestion of 1D gel bands or 2D gel spots.

In-gel digestion was performed using in-house protocols (Benvenuti et al., 2002; Huang et al., 2002) In short in-gel digestion consists of washing, reducing, alkylating, digesting and extracting the peptides from gel bands or plugs. Gel bands were cut into pieces,

[3] http://www.promega.com/tbs/9piv511/9piv511.pdf

approximately 1 mm^3 and placed into 0.65 mL siliconized tubes. Gel plugs were washed and dehydrated by washing several times with 25mM NH_4HCO_3 in 50% ACN.

Gel pieces were dried under vacuum and 10 mM DTT in 25 mM NH_4HCO_3 was added. The reaction was left at 56 °C for 1 hour. After removing the supernatant, an equal volume of 55 mM iodoacetamide 25 mM NH_4HCO_3 was added and the reaction was incubated in the dark for 45 minutes at room temperature.

The supernatant was removed and the gel pieces were washed/dehydrated with 25 mM NH_4HCO_3 in 50% ACN. Gel pieces were fully dried under vacuum and 5 ng/μL trypsin in 25mM NH_4HCO_3 was added just enough to cover the gel pieces. The gel pieces were left rehydrating on ice or at 4°C for 1 hour. Enough 25mM NH_4HCO_3 was added to cover the gel pieces. The reaction was incubated at 37°C for at least 4 hours or overnight.

Peptides were extracted by adding 50 μL 50% ACN / 5% formic acid. The extraction was carried out by vortexing for 30 minutes before collecting the supernatant collected in a separate tube. This process was repeated three times. Sample was dried under vacuum and stored at -80 °C until mass spectrometric analysis.

2.5.2 In-solution digestion of 2D-LC fractions

250 μL of sample was concentrated under vacuum to 50 μL. 10 μL 250 mM NH_4HCO_3 were added together with 10 μL Tris(2-carboxyethyl)phosphine (TCEP) in 25 mM NH_4HCO_3. The vial was overlaid with argon. The reaction was left at 70 °C for 30 minutes whilst being vortexed at 300 rpm. The solution was cooled down to room temperature and 10 μL 200 mM iodoacetamide in 250 mM NH_4HCO_3 were added. The reaction was left at room temperature for 1 hour in the dark.

30 μL were taken out for LC-ESI-MSMS analysis. To approximately 50 μL of solution, 20 ng of trypsin were added. This was incubated at 37 °C for 4 hours at 300 rpm. The solution was stored at -80 °C until further analysis.

2.6 Methanol-Chloroform precipitation of proteins

Protein precipitation was done according to (Wessel and Flugge, 1984). This is a procedure for precipitating proteins from solution, including detergent solutions.

To 0.1 mL of solution 0.4 mL of methanol was added. The solution was mixed rigorously and centrifuged at 9000g for 10 seconds. 0.1 mL of chloroform was added. The solution was mixed rigorously and centrifuged at 9000g for 10 seconds. 0.3 mL of water was added

and mixed vigorously. The mixture was then centrifuged at 9000g for 1 minute. The upper phase was removed. 0.3 mL of methanol was added to the lower phase and interphase with precipitated protein. This mixture was mixed vigorously and centrifuged at 9000g for 2 min to pellet the proteins. Fluid was taken off without disturbing the pellet. The whole mixture was dried under vacuum before being resuspended in standard Laemmli buffer (Laemmli, 1970)

2.7 Preparation of 2D-LC samples for mass spectrometric analysis

To improve the sensitivity and to eliminate the ACN and TFA from the chromatographic elution, 250 µL of each fraction were taken and were concentrated to approximately 50 µL. 2 µL was taken out for MALDI-TOF MS and tandem MS measurement. 1 µL was spotted onto a stainless steel target and 1 µL matrix solution was added. The matrix consisted of 1 part 70% MeOH, 1 part saturated HCCA, 25 fmol μL^{-1} ACTH.

The remaining 48 µL were brought back up to approximately 100 µL with 25 mM ABC. 10 µL 250 mM ABC and 10 µL 20 mM TCEP in 25 mM ABC were added. The solution was overlaid with argon and incubated at 70 °C under shaking (300 rpm) for 30 minutes. The solution was spun down and was allowed to cool down to room temperature before adding 10 µL 200 mM IAM in 25 mM ABC. The solution was incubated at room temperature in the dark for 1 hour.

30 µL were taken out and cleaned up using the ZipTip protocol (see 2.8). This resulted in 10 µL cleaned up sample of which 1 µL was used for MALDI-TOF MS/MS and 1 µL was used for LC-ESI-MS/MS analysis. To approximately 70 µL, 20 ng of trypsin was added. The solution was incubated at 37 °C while shaking (750 rpm) for 4 hours. The solution was concentrated under vacuum to approximately 10 µL and cleaned up using the ZipTip protocol. From the resulting 10 µL, 1 µL was used for MALDI-TOF MS/MS and 1 µL was used for LC-ESI-MS/MS analysis.

2.8 Sample clean up using C18 pre-packed tips

Throughout the work in this thesis, samples were cleaned up using the ZipTip protocol. This name is derived from commercially available tips which are pre-packed with C18 material. The clean-up is based on reversed phase clean up. The tips used were both ZipTip's from Millipore (both normal and micro bedding) and Omix tips from Varian.

Tips from both manufactures where tested using a standard peptide mix and gave very comparable results.

The packing material was wetted three times with 10 µL 50 % ACN using a standard 10 µL pipette. The tip was washed one time with 0.5 % TFA. The sample solution was let through the tip 10 times by slowly aspirating and dispensing without creating air bubbles. The tip was washed two times, with a fresh solution, of 0.5 % TFA. 10 µL 50 % ACN / 0.5 TFA was aspirated and dispensed into a new siliconised tube. The solution was let pipetted up and down the tip another five times.

3 Analysis of *Conus* venom by acrylamide gel electrophoresis based protein separation methods

This chapter is concerned with the analysis of the protein content of several *Conus* venoms. In general research groups only focus on the bio-active peptides. Protein components were analysed by acrylamide gel electrophoresis combined with mass spectrometry and database interrogations. *Conus ventricosus* was analysed by conventional SDS PAGE, *Conus nussatella* by two-dimensional gel electrophoresis (2DE), while *Conus textile* was analysed by both techniques. After separation, selected bands and spots were excised and treated with a protease. Peptides were extracted from the gel and analysed by mass spectrometry. The MS data was searched against the NCBI database.

3.1 Sample origin and preparation.

During the practical work of this thesis, several *Conus* venoms were used. An overview is given in Table 3.1.1. Animals were kept in sea water until they were used for the isolation of venom. The outer shells were crushed using a metal vice. The venom ducts were isolated using ordinary forceps. The venom was released into ddH_2O by gently pressing with a pestle. The solution was spun down to remove cellular debris. This process was repeated once more with the supernatant. The resulting supernatant was collected, lyophilised and stored at -80 °C until further use. Venoms were supplied by Dr. Michael Fainzilber from the Weizmann Institute of Science, Rehovot Israel. The venom from *Conus nussatella, Conus textile* and *Conus Arenatus* was supplied as crude venom and lyophilised powder from the void volume peak from a Sephadex G-50, hence supposedly having components above 20 kDa. *Conus ventricosus* and *Conus nussatella* were supplied as crude venom, prepared as described above.

Table 3.1.1 Overview of *Conus* venoms used.

A multitude of *Conus* venoms were used throughout the course of this project, an overview is given in this table. The column gives information about the name and the abbreviation, the location where the species is found and on which organism it preys. Vermivorous prey on worms, Piscivorous on fish and Molluscivorous on snails.

Conus	Abbreviation	Prey	Location
ventricosus	C.Ventri	Vermivorous	Mediterranean Sea
textile	C.Tex	Piscivorous	Indo-Pacific
nussatella	C.Nuss	Molluscivorous	Indo-Pacific
arenatus	C.Ar	Vermivorous	Red Sea

The venoms of *Conus ventricosus* and *Conus textile* were pre-separated before being analysed by 1D SDS-PAGE using a size exclusion column. The Tricorn Superdex peptide 10/300GL column (Amersham Biosciences) had an optimal separating range from 100 to 7,000 Da. By using this column large proteins were separated from the peptide and small molecule part of the venom.

The column was calibrated using a mixture of proteins and peptides described in Table 3.1.2. The calibration mixture functioned to test the co-elution of proteins and to calibrate the size column.

Figure 3.1.1 shows an overlay chromatogram of the calibration mixture and a test sample. Independent injection of Transferrin and BSA (data not shown) illustrated that these two proteins have identical elution times. Therefore proteins with a molecular weight greater then 66 kDa co-elute (Peak A). The remaining components were separated (Peak B through E).

Table 3.1.2 Constituents of size column calibration mixture.

This table shows the constituents, approximate molecular weight, amount loaded onto the column and the elution time of the mixture used to calibrate the size column. The mixture was designed the test the co-elution of (large) protein and to calibrate the optimal separation range. The corresponding chromatogram is shown in Figure 3.1.1.

Peak	Component	Aprox. Molecular Weight	Loaded on Column (µg)	Elution time (min)
A	Transferrin	85 kDa	5.4	35.8
	Bovine Serum Albumin	66 kDa	2.9	
B	Carbonic Anhydrase	29 kDa	4.9	40.3
C	Cytochrome C	12.4 kDa	1.5	45.7
D	Insulin	4749.5 Da	7.8	59.5
E	Substance P fragment	1191.5 Da	2.4	77.9

Trypsin digested fetuin (bovine) was prepared according to Medzihradszky *et al.* (Medzihradszky et al., 1994). Briefly, bovine fetuin was dissolved in 6M guanidine HCl/300 mM N-ethylmorpholine acetate solution (pH 8.3), reduced with DTT and alkylated with vinylpyridine. After dialysis against ammonium carbonate, the protein was digested with the serine protease trypsin. Roughly 100 pmol was loaded and used to check the performance of the column. Fractions were taken every 1.5 mL, indicated in blue on the X-axis. Masses were observed from each fraction with standard LC-ESI tandem MS. The first of the two major peaks in between D and E contained large glyco-peptides observed by Medzihradszky *et al.* (Medzihradszky et al., 1994). The second of the two peaks contained the 'normal' tryptic peptides in the mass range determined by the calibration.

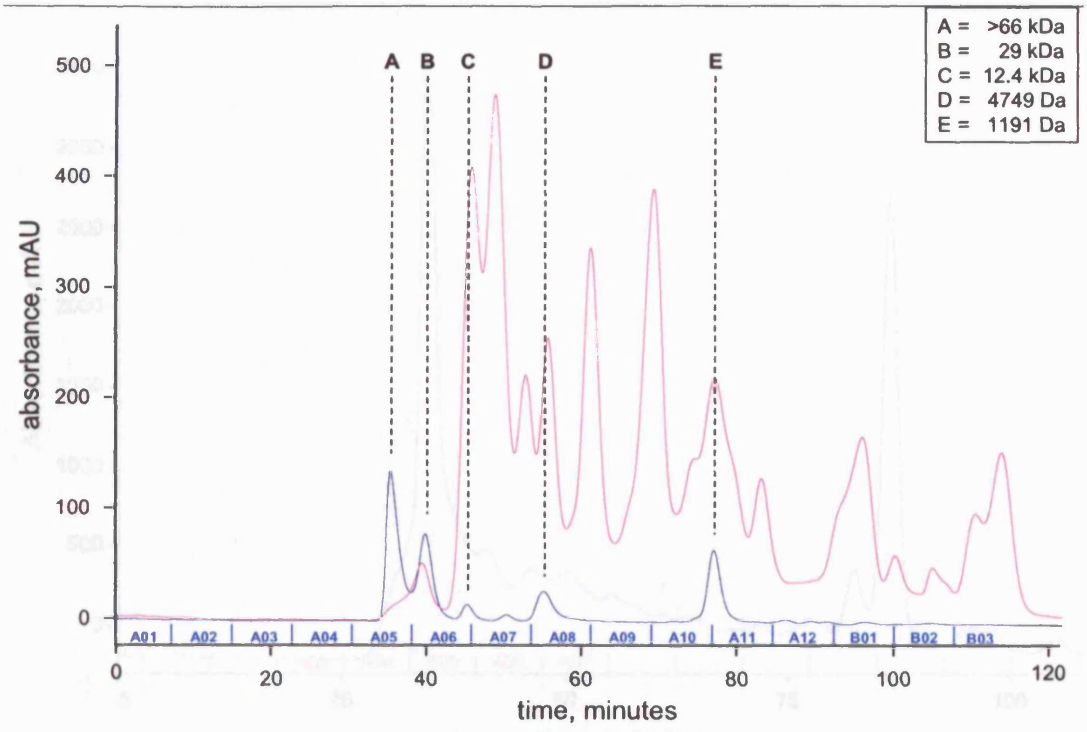


Figure 3.1.1 Size column calibration and testing.

This figure shows an overlay of chromatograms of the UV absorbance (214 nm) of size separated calibration mixture (blue) and the chromatogram of a trypsin digest of Bovine Fetuin (magenta). (A) is the co-elution of Transferrin (85 kDa) and Bovine Serum Albumin (66 kDa) (E et al.) is Carbonic Anhydrase (29 kDa) (E et al.) is Cytochrome C (12.4 kDa) (D) is Insulin (4749 Da) and (E et al.) is Substance P fragment (1191.5 Da). Fractions from the fetuin digest run were taken every 1.5 mL (blue).

3.2 1D-SDS PAGE separation and analysis of Conus ventricosus venom.

Conus ventricosus venom was separated on a size exclusion column. To this, 6.3 µg lyophilised *Conus ventricosus* venom was mixed with 175 µL ASW. It was left shaking for 1.5 hours (750 rpm) at 4 °C. The sample was spun down and the supernatant was transferred to a clean tube. The supernatant had an orange/yellow colour. Another 175 µL ASW was added to the pellet and was left shaking for 1 hour (750 rpm) at 4 °C. This process was repeated one more time. The supernatants were combined to give approximately 520 µL solubilised venom. 90 µL was used to prepare protein samples. The UV absorbance chromatogram (214 nm) of this size separation is shown in Figure 3.2.1.

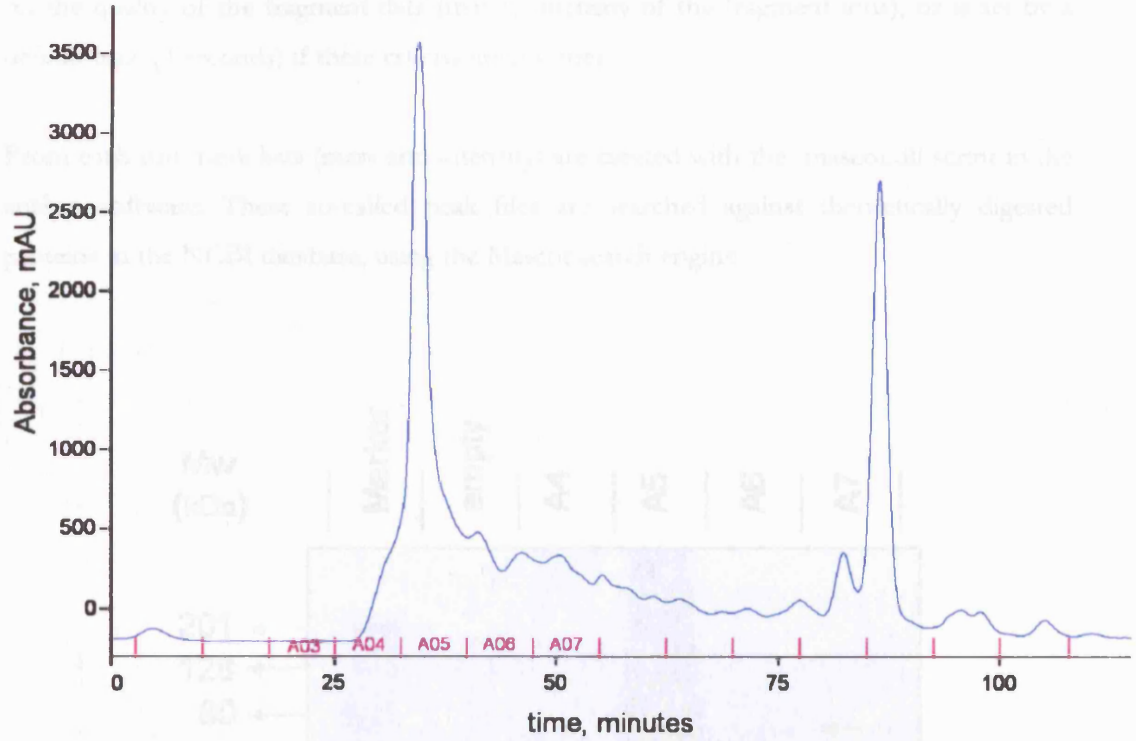


Figure 3.2.1 Elution profile from size separation of *Conus ventricosus* venom.

The figure shows the absorbance over time of the size separation of *Conus ventricosus* venom. The blue line is the absorbance at 214 nm. Fractions were taken every 1.5 mL (magenta). Fraction A04, A05, A06 and A07 were further analysed by 1D-SDS-PAGE.

Proteins from fraction A04 through A07 were precipitated to dispose of salts from the elution buffer (150 mM NaCl, 10 mM Tris). The precipitation was done according to Wessel *et al.* (Wessel and Flugge, 1984). The individual fractions were loaded onto a standard SDS-PAGE gel, with a 10-20% acrylamide gradient. See Figure 3.2.2 for the colloidal coomassie blue stained gel. The very intense blue colour indicated that all proteins co-eluted in fraction A05. Faint coloured bands were visible in lane A04 and A06 (indicated by the black arrows), while any form of stained protein was absent in A07.

Lane A05 was cut in 20 equidistant slices. Together with the bands indicated in lane A04 and A06, these were subjected to in-gel digestion. The in-gel digestion was performed using in an house protocol (Benvenuti et al., 2002; Huang et al., 2002).

Each of the in-gel digestions was subjected to online LC-ESI tandem mass spectrometry. Data was acquired in a data dependent matter, where ions above certain intensity are automatically selected for fragmentation. The duration of this fragmentation is dependent

on the quality of the fragment data (mainly intensity of the fragment ions), or is set by a default time (3 seconds) if these criteria are not met.

From each run, peak lists (mass and intensity) are created with the mascot.dll script in the analyst software. These so-called peak files are searched against theoretically digested proteins in the NCBI database, using the Mascot search engine.

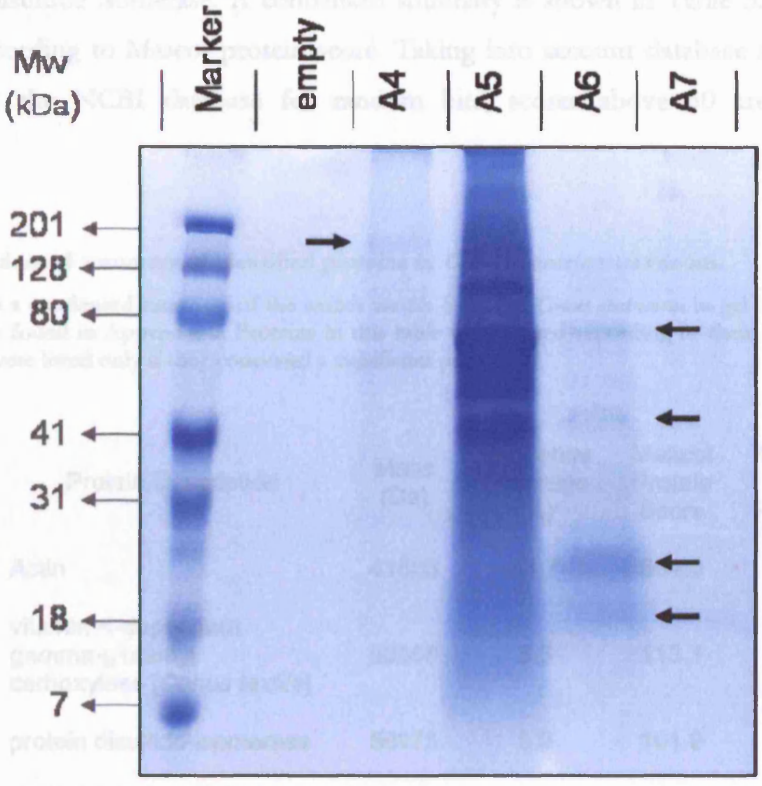


Figure 3.2.2 1D SDS PAGE of size exclusion fractions from *Conus ventricosus*.

The size exclusion fractions from the size separation of *Conus ventricosus* (Figure 3.2.1) were desalted by protein precipitation, concentrated and loaded on different lanes on a 10-20% acrylamide gel. The proteins were visualised using a colloidal coomassie blue staining. Bands where excised from lane A4 and A6, indicated by the arrows. Lane A5 was divided in 20 equidistant bands. Proteins in the bands were reduced, alkylated, digested with trypsin and extracted. Samples were analysed by LC-ESI-MS\MS.

A complete list of proteins identified by mass spectrometry can be found in Appendix 2. This list is filtered for human keratin and the autolysis products of modified porcine trypsin (Promega). The protein hits displayed from an MS/MS run contained at least one peptide with a Mascot peptide score greater than the 5% significance threshold for the NCBI database.

The proteins that were identified were, with the exception of few, from other organisms. However, *Conus* species were underrepresented in databases. To date only 1087 protein entries were present for all the 500 known species. Many of the protein entries are predicted precursor conotoxins which fall below the detection limit of conventional SDS-PAGE i.e. with a mass smaller than 7 kDa.

Therefore, proteins were identified by homologous peptides. Hence, it was not surprising that many of the protein hits are well known, abundant household proteins, such as actin and protein disulfide isomerase. A condensed summary is shown in Table 3.2.1. Proteins are sorted according to Mascot protein score. Taking into account database size and 5 % threshold for the NCBI database for random hits, scores above 50 are considered significant.

Table 3.2.1 condensed summary of identified proteins in *Conus ventricosus* venom.

This table shows a condensed summary of the search results from the *Conus ventricosus* in-gel digests. The full summary can be found in Appendix 2. Proteins in this table were sorted according to their Mascot protein score. Proteins were listed only if they contained a significant peptide

Accession Number	Protein Description	Mass (Da)	Sequence Coverage (%)	Mascot Protein Score	Number of Peptides	pI
3182894	Actin	41683	44.7	564.0	26	5.30
18652793	vitamin K-dependent gamma-glutamyl carboxylase [*Conus textile*]	93858	3.5	113.1	3	5.37
19880309	protein disulfide-isomerase	56975	3.9	101.9	4	4.78
112696	14-3-3 protein zeta/delta (Protein kinase C inhibitor protein-1) (KCIP-1)	27839	13.1	81.1	4	4.77
3318722	Chain E, Leech-Derived Tryptase Inhibitor	23457	12.6	76.5	2	8.26
138531	Vimentins 1 and 2	52812	4.1	71.9	3	5.16
138532	Vimentin 4	53464	4.1	71.9	3	5.08
33591156	thioredoxin peroxidase [Ixodes ricinus]	19134	10.7	71.3	4	6.93
5902072	serine proteinase inhibitor, clade B (ovalbumin), member 3; squamous cell carcinoma a	44537	5.1	70.0	2	6.35

44829400	peroxiredoxin 1 [Ixodes ricinus]	18977	10.6	69.6	4	11.80
42523784	thiosulfate sulfurtransferase [Bdellovibrio bacteriovorus HD100]	31343	4.3	68.6	1	9.45
16945685	disulfide isomerase [Ostertagia ostertagi]	54972	2.8	61.5	3	4.93
204499	glutathione S-transferase Y-b subunit (EC 2.5.1.18)	21871	4.8	60.1	1	7.82
13021648	peptidyl-prolyl cis-trans isomerase B [Xenopus laevis]	11343	10.6	52.3	3	9.36

3.3 1D-SDS PAGE separation and analysis of Conus textile venom.

Conus textile venom was separated on a size exclusion column as in section 3.2. 90 µL were used to prepare protein samples. The UV absorbance chromatogram (214 nm) of this size separation is shown in Figure 3.3.1

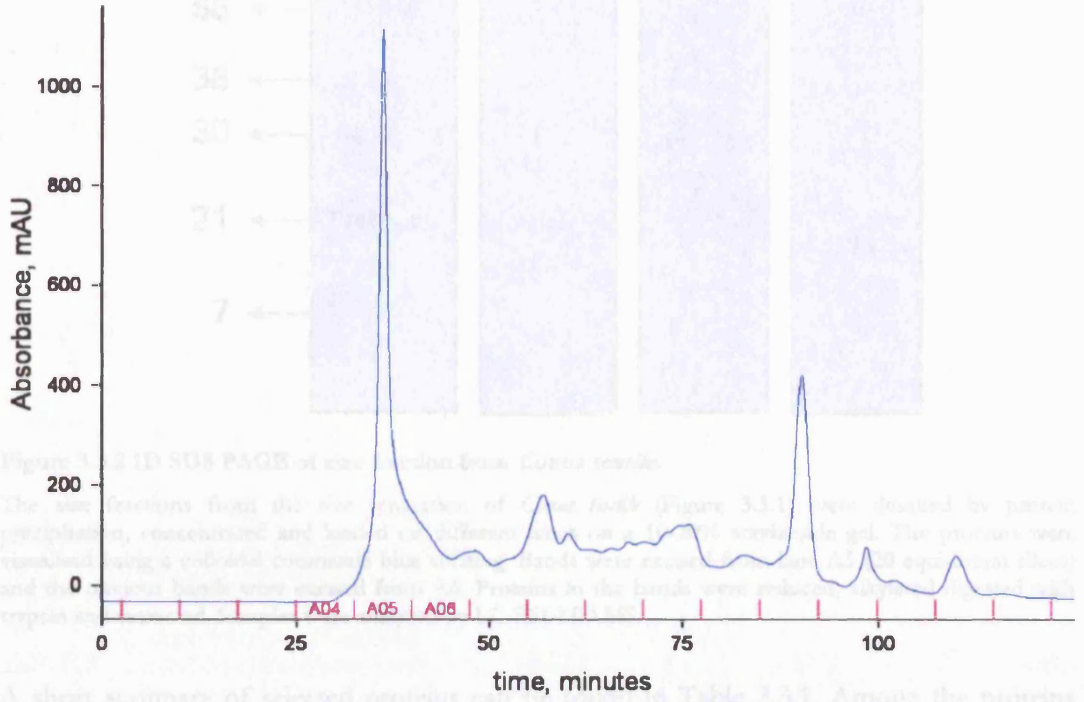


Figure 3.3.1 Elution profile from size separation of *Conus textile* venom.

The figure shows the absorbance over time of the size separation of *Conus* Texile venom. The blue line is the absorbance at 214 nm. Fractions were taken every 1.5 mL (magenta). Fraction A04, A05 and A06 were further analysed by 1D-SDS-PAGE.

In a similar fashion to *Conus ventricosus*, fraction A04, A05 and A06 were desalted by protein precipitation, according to Wessel *et al.* (Wessel and Flugge, 1984). The proteins were then dissolved in 1D sample buffer and run on a standard 10-20% gradient gel. Proteins were stained using colloidal coomassie blue. The resulting staining pattern is shown in Figure 3.3.2

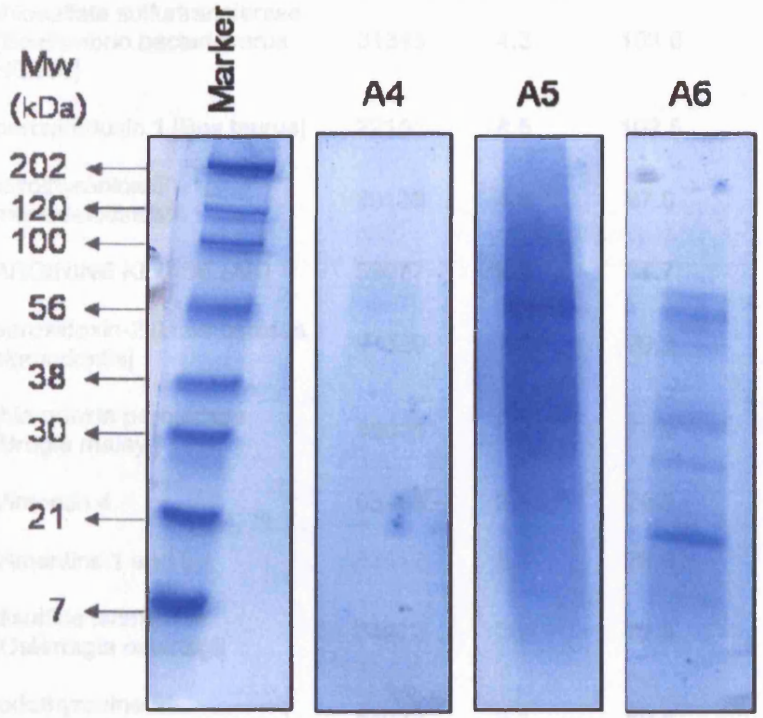


Figure 3.3.2 1D SDS PAGE of size fraction from *Conus textile*.

The size fractions from the size separation of *Conus textile* (Figure 3.3.1) were desalted by protein precipitation, concentrated and loaded on different lanes on a 10-20% acrylamide gel. The proteins were visualised using a colloidal coomassie blue staining. Bands were excised from lane A5 (20 equidistant slices) and the obvious bands were excised from A6. Proteins in the bands were reduced, alkylated digested with trypsin and extracted. Samples were analysed by LC-ESI-MS\MS.

A short summary of selected proteins can be found in Table 3.3.1. Among the proteins found was a previously reported (in *Conus textile*) carboxylase (Stanley et al., 1997). Protein disulfide isomerase, a protein involved with protein folding and formation of disulfide bridges was also found. Arginine kinase (AK) belongs to a class of kinases that play a role in the maintenance of ATP levels by the phosphorylation of the so called "phosphagens". This also has been previously identified.

Table 3.3.1 Condensed summary of proteins identified from *Conus textile* venom.

This table shows a condensed summary of the search results from the *Conus textile* in-gel digest. The full summary can be found in Appendix 3. Proteins in this table were sorted according to Mascot protein score. Only proteins with a significant peptide were listed.

Accession Number	Protein Description	Mass (Da)	Sequence Coverage	Mascot Protein Score	Number of Peptides	pI
10834855	Actin	22879	40.1	343.6	12	4.92
23463261	vitamin K-dependent carboxylase [*Conus textile*]	93905	7.8	272.6	6	5.37

42523784	thiosulfate sulfurtransferase [Bdellovibrio bacteriovorus HD100]	31343	4.3	103.6	1	9.45
27806081	peroxiredoxin 1 [Bos taurus]	22151	8.5	102.5	2	8.81
202547	iodothyronine 5"-monodeiodinase	30136	4.9	97.0	4	5.09
3183057	ARGININE KINASE (AK)	39077	9.5	87.7	3	7.12
9965598	peroxidoxin-2 [Litomosoides sigmodontis]	21320	7.8	79.2	2	6.41
12751382	thioredoxin peroxidase [Brugia malayi]	18005	9.2	79.2	2	6.28
138532	Vimentin 4	53464	3.5	76.8	2	5.08
138531	Vimentins 1 and 2	52812	3.5	76.8	2	5.16
16945685	disulfide isomerase [Ostertagia ostertagi]	54972	2.8	72.9	3	4.93
202549	iodothyronine 5" monodeiodinase	54033	3.1	69.4	2	4.87
27806501	procollagen-proline, 2-oxoglutarate 4-dioxygenase [proline 4-hydroxylase], beta polypeptide [protei	57230	2.9	69.4	2	4.80
3551774	alpha-glycerophosphate oxidase [Streptococcus pneumoniae]	66751	2.1	67.2	2	4.98
37362272	2-peptidylprolyl isomerase A [Danio rerio]	17390	9.1	63.5	2	8.87
12230403	Probable proline iminopeptidase (PIP) (Prolyl aminopeptidase) (PAP)	36452	2.2	53.1	1	5.81

3.4 Analysis of Conus venoms by two-dimensional gel electrophoresis

This paragraph describes the analysis of the following *Conus* venoms (i) *C. arenatus* (ii) *C. nussatella* (iii) *C. textile.* These samples were kindly provided by Dr. Mike Fainzilber from the Weizmann Institute of Science, Rehovot Israel.

3.4.1 Sample solubilisation and labelling of proteins with CyeDye™

Lyophilised *Conus* venom was dissolved in 2DE sample buffer (8 M, 4% w/v CHAPS, 0.5% (v/v) NP-40, 10 mM Tris-HCl). Protein concentration was determined using

Coomassie Protein Assay Reagent and a BSA standard curve with 3 replicate assays performed for each sample. Samples were adjusted to the same protein concentration using 2D buffer. For protein concentration of venom stock solutions see Table 3.4.1

Table 3.4.1 Protein concentration of venom solutions.

Lyophilised venom was weight and dissolved in 2DE sample buffer. Protein concentrations were determined against a BSA standard curve (shown in Figure 3.4.1). The accuracy of the determination is shown by the 95% prediction interval (PI). CV stands for Crude Venom and V0 for the void volume of a size exclusion run.

Venom	weight (mg)	concentration (mg/ml)	± (95 % PI)
(1) C.Ar – CV	9.3	30.8	3.7
(2) C.Nuss – CV	7.8	19.8	2.7
(3) C.Ar - V0	7.2	40.0	4.6
(4) C.Tex - V0	7.3	20.3	2.8

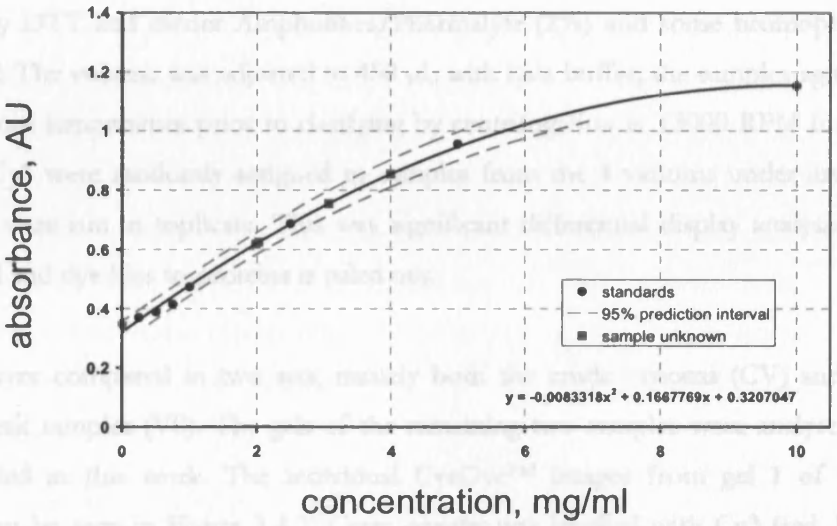


Figure 3.4.1 Standard curve for protein determination in *Conus* venoms.

A regression analysis was performed for the protein concentration determination, using an in house Excel spreadsheet. The accuracy of the measurement is modelled by the prediction intervals (p=0.05).

A sample amount of 150 µg was aliquoted into tubes for CyDye™ labelling. Equal amounts of protein from each sample were mixed to create an internal standard which was labelled with Cy2. Enough material was provided to run the internal standard on every gel. Cy3- and Cy5-labelled samples were mixed appropriately (See Table 3.4.2). The samples

were run in triplicate and labeled both with Cy3 and Cy5. This was to avoid possible dye bias was and to enable significant differential analysis could be performed.

Table 3.4.2 Sample labelleling scheme for CyDye™ experiment.

This table shows scheme used labelling the venom camples. Crude venom (CV) and the void peak (V0) from a sephadex G-50 elution were used.

Gel number	Cy3	Cy5
1	*C. arenatus* **CV**	*C. nussatella* **CV**
2	*C. arenatus* **CV**	*C. arenatus* **V0**
3	C. *textile* **V0**	*C. arenatus* **CV**
4	*C. nussatella* **CV**	*C. arenatus* **V0**
5	*C. nussatella* **CV**	C. *textile* **V0**
6	*C. arenatus* **V0**	C. *textile* **V0**

A suitable amount (150 µg) of the Cy2-labelled pool was added to each gel. Samples were reduced by DTT and carrier Ampholines/Pharmalyte (2%) and some bromophenol blue was added. The volume was adjusted to 450 µL with lysis buffer, the samples agitated for 2 min. at room temperature prior to clarifying by centrifugation at 13000 RPM for 10 mins. Cy3 and Cy5 were randomly assigned to samples from the 4 venoms under investigation and these were run in triplicate. This way significant differential display analysis could be performed and dye bias to proteins is ruled out.

Samples were compared in two sets, namely both the crude venoms (CV) and the void volume peak samples (V0). The gels of the remaining two samples were analysed, but are not included in this work. The individual CyeDye™ images from gel 1 of the crude venoms can be seen in Figure 3.4.2. *Conus arenatus* was labelled with Cy3 (red, left panel) and *Conus nussatella* with Cy5 (blue, right panel). The protein expression profile of the two species is completely different. This is demonstrated in Figure 3.4.3. This figure displays the overlay image of the two CyDye images. In a more traditional differential 2DE analysis, for instance investigating the influence of treatment of a cell line with a compound of interest, usually the majority of spot (proteins) overlap (Hoving et al., 2002).

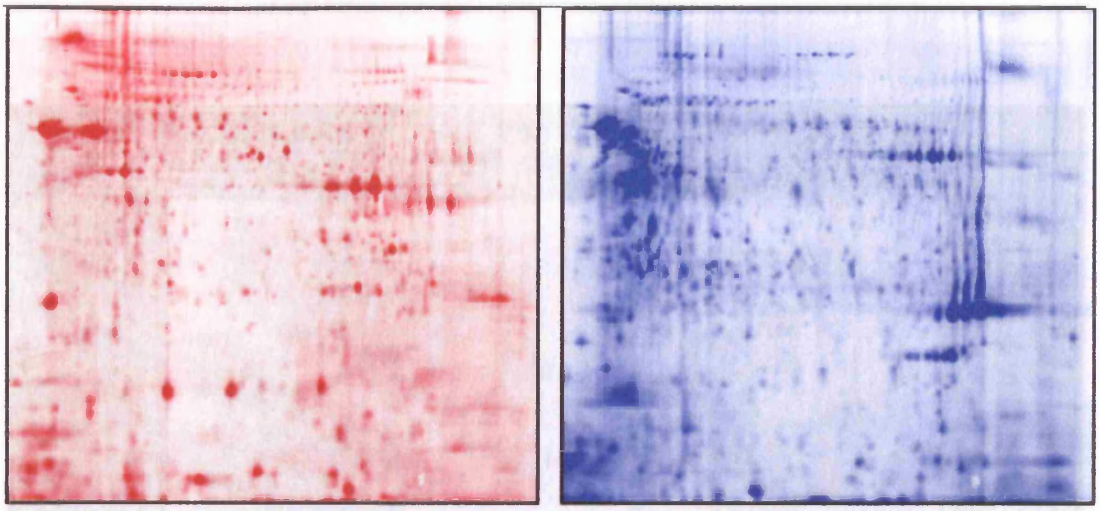


Figure 3.4.2 2DE CyeDye images from *Conus arenatus* and *nussatella*.

This figure shows the Cy3 (left) and Cy5 (right) image from gel 1 (see table Table 3.4.2). These were the venoms of *Conus arenatus* and *Conus nussatella* respectively.

Identical spots are represented by black spots, indicating no change in protein expression, or modification, between the two samples. Some of these unaltered spots can be seen upon the comparison of the two venoms in Figure 3.4.3. The images generated were saturated and analysed using DeCyder™ software. Briefly the difference in-gel analysis (DIA) module was used to define spot boundaries and to measure normalized spot volumes. The biological variation analysis (BVA) mode of DeCyder™ was used to match all pairwise image comparisons from DIA for cross-gel statistical analysis. Spots displaying a ≥ 5 average fold increase or decrease in abundance (p value < 0.05) were selected for identifications. The excised spots were circled in yellow. These were digested in-gel and analysed by LC-ESI-MS/MS. Tandem MS data was then searched against the NCBI protein database.

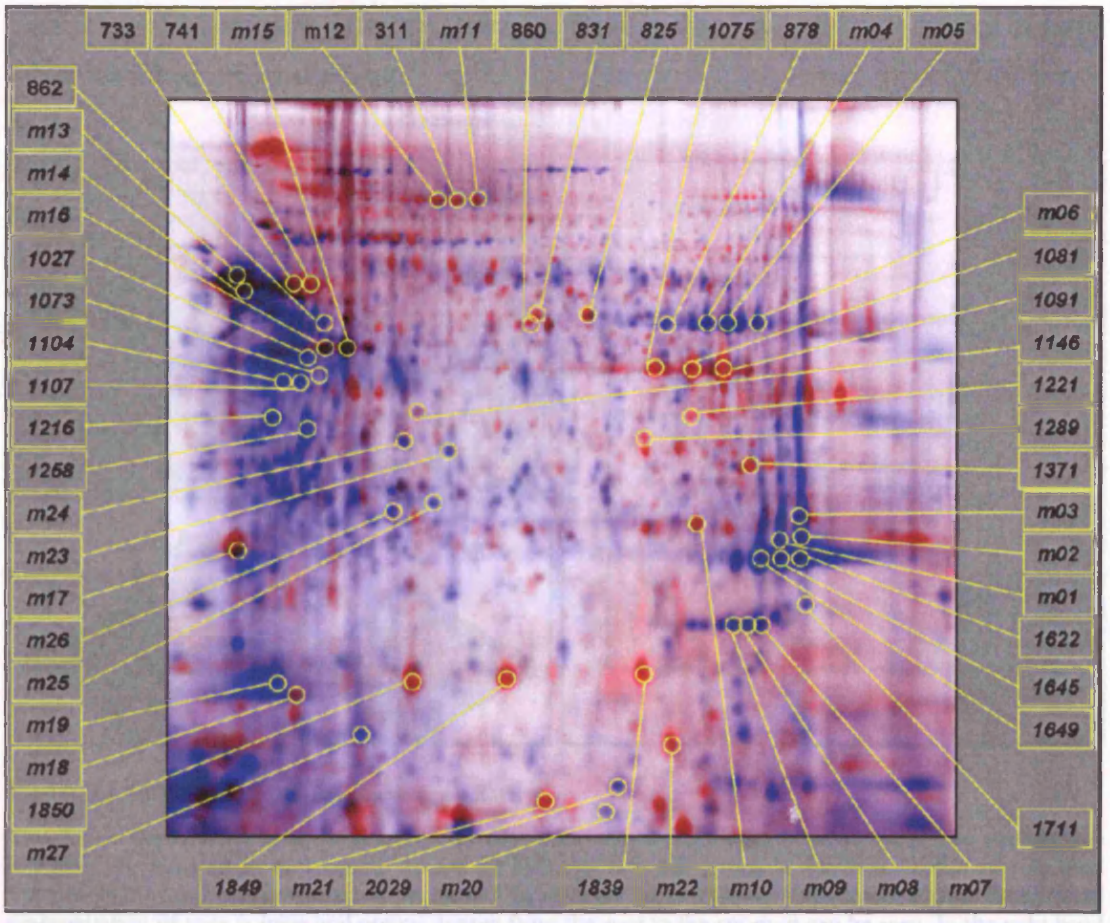


Figure 3.4.3 2DE overlay image of *Conus arenatus* and *nussatella*.

This figure shows the overlay of the images from Figure 3.4.2. Spots that appear in black are common on both gels in the same amount. Red spots mean that the particular protein is more abundant for the Cy3 labelled samples, vice versa for the blue spots. Spots circled were excised, digested and analysed by mass spectrometry. Numbers starting with 'm' where selected manually to be picked by the picking robot. The other spot numbers were generated by the DeCyder™ analysis.

Three examples of spot analyses are given in Figure 3.4.4, Figure 3.4.5 and Figure 3.4.6. In each of these figures, three different panels can be seen. Top left, a zoom-in of the actual gel image. Bottom left a 3D graphical representation of the spot volume and density. On the right hand side a graph with the relative abundance of that particular spot across the four samples investigated. Venoms 1 through 4 correspond to the numbers in Table 3.4.1.

In Figure 3.4.4 the DeCyder output for spot 1081 is shown. This is a spot, or feature, only seen in the crude venom sample of *Conus arenatus* (red Cy3 image from Figure 3.4.2). This spot was actually saturated, hence the flattened top of the densitogram. This protein was identified as Arginine Kinase (accession number NCBI 3183057). See also Table 3.4.3 with

a summary of the proteins identified from gel 1 (*Conus arenatus* and *Conus nussatella*). This protein was also identified during the 1D-PAGE experiments on the crude venoms of *Conus ventricosus* and *textile*. However, from the spots analysed for the crude venom of *Conus nussatella*, this protein was absent.

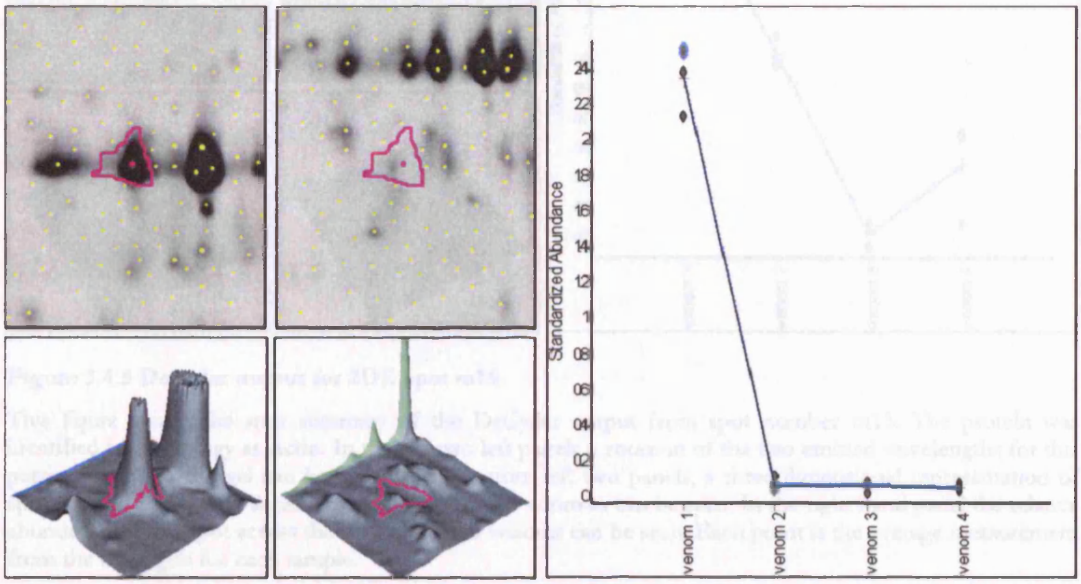


Figure 3.4.4 DeCyder output for 2DE spot 1081.

This figure displays the spot summary of the DeCyder output from spot number 1081. The protein was identified by homology, as Arginine kinase. In the top two left panels a zoom-in of the two emitted wavelengths for this particular part of the gel can be seen. In the bottom left two panels, a three dimensional representation of spot volume and optical density from the spot in the zoom-in can be seen. In the right hand panel the relative abundance of this spot across the four examined venoms can be seen. Each point is the average measurements from the three gels for each sample.

Figure 3.4.5 is an example of a spot that has similar spot intensity for the first two venoms (black spots from Figure 3.4.3). This spot was identified as Actin (accessions number 224306). A further three spots in both venoms were identified as Actin.

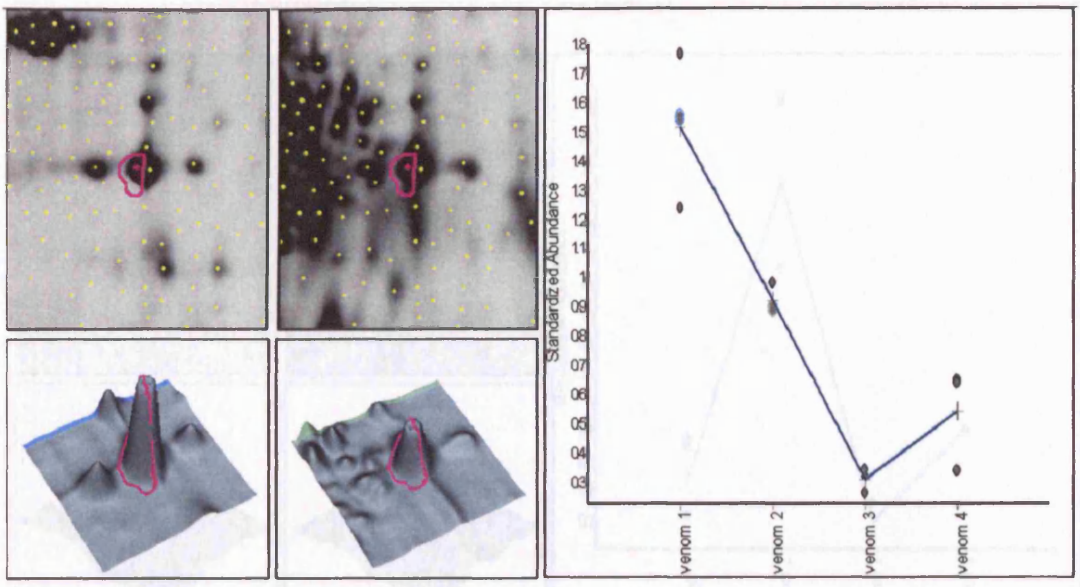


Figure 3.4.5 Decyder output for 2DE spot m15.

This figure shows the spot summary of the DeCyder output from spot number m15. The protein was identified by homology as Actin. In the top two left panels a zoom-in of the two emitted wavelengths for this particular part of the gel can be seen. In the bottom left two panels, a three dimensional representation of spot volume and optical density from the spot in the zoom-in can be seen. In the right hand panel the relative abundance of this spot across the four examined venoms can be seen. Each point is the average measurement from the three gels for each sample.

Figure 3.4.6 is an example of a protein differentially expressed in the venom of *Conus nussatella*. Like the previous figure, also this spot was identified as Actin (Accession number 32816054).

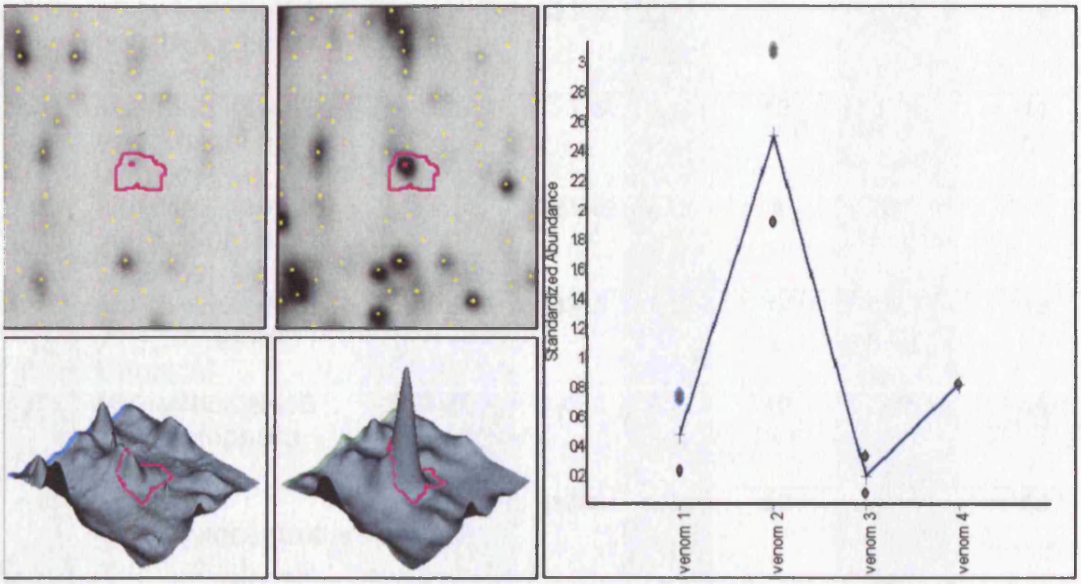


Figure 3.4.6 Decyder output for 2DE spot m24

This figure displays the spot summary of the DeCyder output from spot number m24. The protein was identified by homology, as Actin. In the top two left panels a zoom-in of the two emitted wavelengths for this particular part of the gel can be seen. In the bottom left two panels, a three dimensional representation of spot volume and optical density from the spot in the zoom-in can be seen. In the right hand panel the relative abundance of this spot across the four examined venoms can be seen. Each point is the average measurement from the three gels for each sample.

The overall success rate of identifying proteins was not very high. Out of the 56 spots picked for identification, one spot was lost by the picking robot and from the remaining samples, 12 spots were identified with four unique proteins. An overview of these results is displayed in Table 3.4.3

Table 3.4.3 Summary of identified proteins in *Conus* venom by 2DE.

The results in this table represent a summary of the proteins identified from the selected 2DE gel spots. Besides protein name and accession number, the molecular weight (Mr) and calculated iso-electric point (pI) are given. In the last three columns, the number of peptides used for identifying the protein by tandem mass spectrometry, the Mascot protein score and sequence coverage are given.

Spot	Protein Name	Accesion Number	Mr	pI	number of peptide	Mascot Protein Score	Sequence coverage (%)
741	alpha-tubulin 3 [Ciona intestinalis]	21667231	44564	5.85	3	124	13
831	enolase 2 [Pycnococcus provasolii]	15667700	37659	4.58	3	99	4
1075	ARGININE KINASE (AK) [Liolophura	3183057	39305	7.12	8	188	14

	japonica]						
1081	ARGININE KINASE (AK) [Liolophura japonica]	3183057	39305	7.12	14	218	14
1091	ARGININE KINASE (AK) [Liolophura japonica]	3183057	39305	7.12	18	275	14
1221	ARGININE KINASE (AK) [Liolophura japonica]	3183057	39305	7.12	4	128	7
1289	ARGININE KINASE (AK) [Liolophura japonica]	3183057	39305	7.12	5	163	13
1371	ARGININE KINASE (AK) [Liolophura japonica]	3183057	39305	7.12	10	275	14
m15	Actin [Strongylocentrotus purpuratus]	224306	41798	5.30	50	616	52
m16	Actin [Strongylocentrotus purpuratus]	224306	41798	5.30	25	369	52
m24	actin [Chlamys farreri]	32816054	42116	5.29	15	475	33
m25	ACTIN 3 [Echinococcus granulosus]	543768	35027	5.25	17	405	41

4 Liquid Chromatography MALDI-MS/MS analysis of *Conus ventricosus* Venom.

Liquid chromatography MALDI tandem mass spectrometry, in short LC-MALDI-MS/MS is being used frequently now as a technique to analyse complex biological mixtures (Chen et al., 2005; Floyd et al., 1999). Pilot experiments carried out in the UCSF mass spectrometry facility, showed that conotoxins can be detected in the crude venom of *Conus ventricosus*. In this particular experiment, *Conus ventricosus* venom was solubilised in ammonium bicarbonate (ABC) and separated in two rounds of reversed phase separation. Fractions from the first dimension were re-injected onto the column, only substituting the mobile phase acetonitrile with iso-propanol (2-propanol) to generate a semi orthogonal separation. This last run was directly spotted onto a MALDI target and analysed by mass spectrometry. Results showed motives of a number of so called, contryphans, which are single disulfide linked conotoxins (Jacobsen et al., 1998; Jimenez et al., 1996; Massilia et al., 2001).

4.1 First dimension size separation of Conus ventricosus venom

Conus ventricosus venom, from a different batch was supplied by Dr. Michael Fainzilber. This was solubilised in 25 mM ABC pH 7.8. This buffer was also used as mobile phase for the size exclusion separation of the venom. In this way fractions were in the correct medium for reduction by dithiothreitol DTT, alkylation by iodoacetamide (IAM) and digestion by trypsin.

Figure 4.1.1 shows the scan of a paper trace chromatogram of the size exclusion separation from *Conus ventricosus* venom. For this separation a TSKgel Super SW2000 (Tosoh Bioscience) column was used. Fractions were collected manually after each major peak, as indicated by the red numbers in the figure.

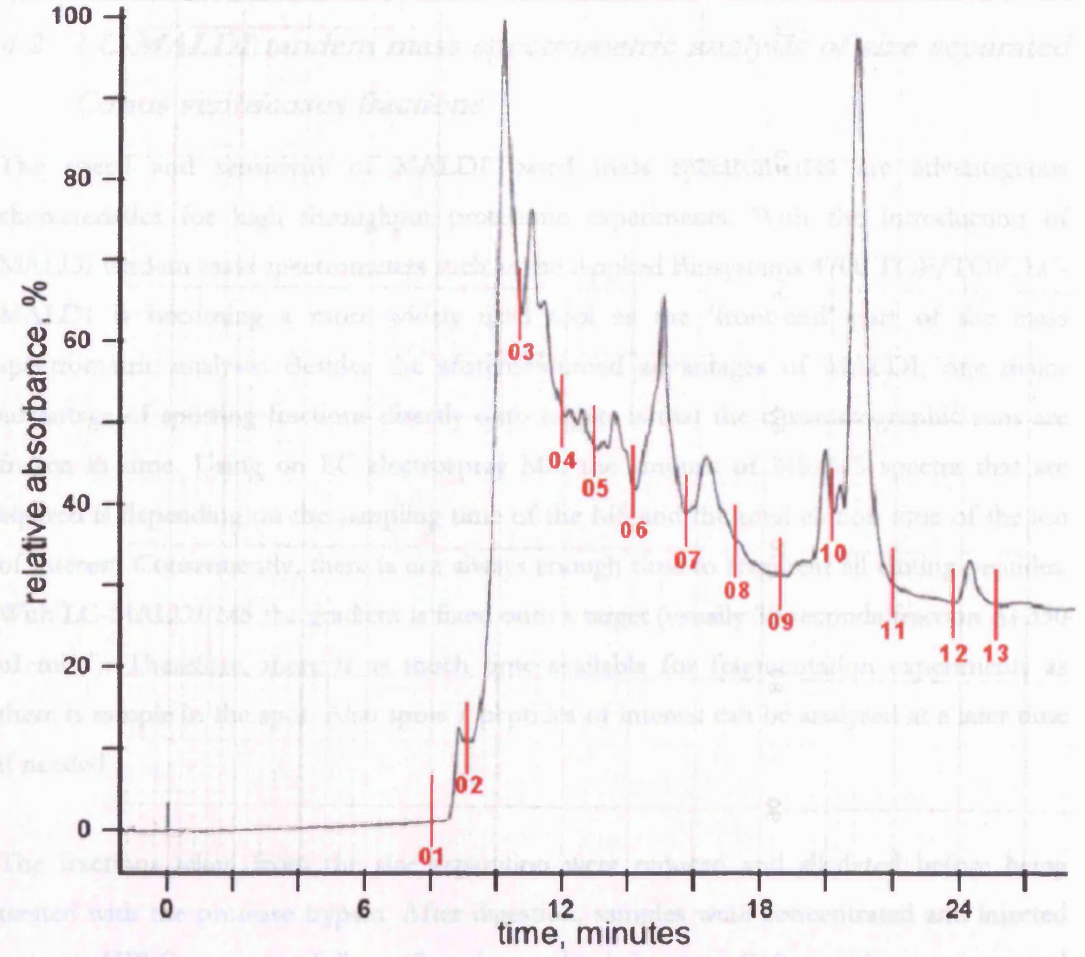


Figure 4.1.1 First dimension separation of *Conus ventricosus* venom.

This figure shows the chromatogram of the absorbance at 214 nm of size separated *Conus ventricosus* venom. The venom was solubilised in the mobile phase ABC at a flow rate of 0.5 mL·min^{-1}. Fractions were taken manually, indicated by the red numbers. Proteins and peptides were reduced directly by DTT. Subsequently after reduction fractions were alkylated and incubated with trypsin. The resulting peptides were analysed by LC-MALDI.

4.2 LC-MALDI tandem mass spectrometric analysis of size separated Conus ventricosus fractions

The speed and sensitivity of MALDI based mass spectrometers are advantageous characteristics for high throughput proteomic experiments. With the introduction of MALDI tandem mass spectrometers such as the Applied Biosystems 4700 TOF/TOF, LC-MALDI is becoming a more widely used tool as the 'front-end' part of the mass spectrometric analysis. Besides the aforementioned advantages of MALDI, one major advantage of spotting fractions directly onto targets is that the chromatographic runs are frozen in time. Using on LC electrospray MS, the amount of MS/MS spectra that are aquired is depending on the sampling time of the MS and the total elution time of the ion of interest. Consequently, there is not always enough time to fragment all eluting peptides. With LC-MALDI-MS the gradient is fixed onto a target (usually 30 seconds fraction At 350 $nL \cdot min^{-1}$). Therefore, there is as much time available for fragmentation experiments as there is sample in the spot. Also spots / peptides of interest can be analysed at a later time if needed.

The fractions taken from the size separation were reduced and alkylated before being treated with the protease trypsin. After digestion, samples were concentrated and injected onto an HPLC system as follows. Samples are loaded onto a C18 peptide trap (i.e. guard column) which is on the place of the sample loop. At this point the samples are washed to remove salts from the sample itself and those introduced by solubilisation and reducing and alkylation. The eluent of the column is directed to waste. After the rinse, the guard column is switched online with the analytical column and peptides were eluted onto the main column. Peptides are separated in a water / organic buffer system. Figure 4.2.1 shows the UV absorbance at 214 nm of the chromatographic separation of each size fraction. The numbers correspond to the fractions in Figure 4.1.1. Fractions were spotted directly onto a stainless steel target. Just before elution onto the target, the eluent is mixed online (t-junction) with alpha cyano 4-hydroxy-cinnamic acid matrix and an internal standard. The internal standard (IS), Adrenocorticotropic Hormone clip 18-38 (ACTH) was at a concentration of approximately 4 fmol per spot. The IS functioned as internal calibration peptide. The mass spectrum is calibrated on the fly automatically by the controlling software of the AB 4700 mass spectrometer.

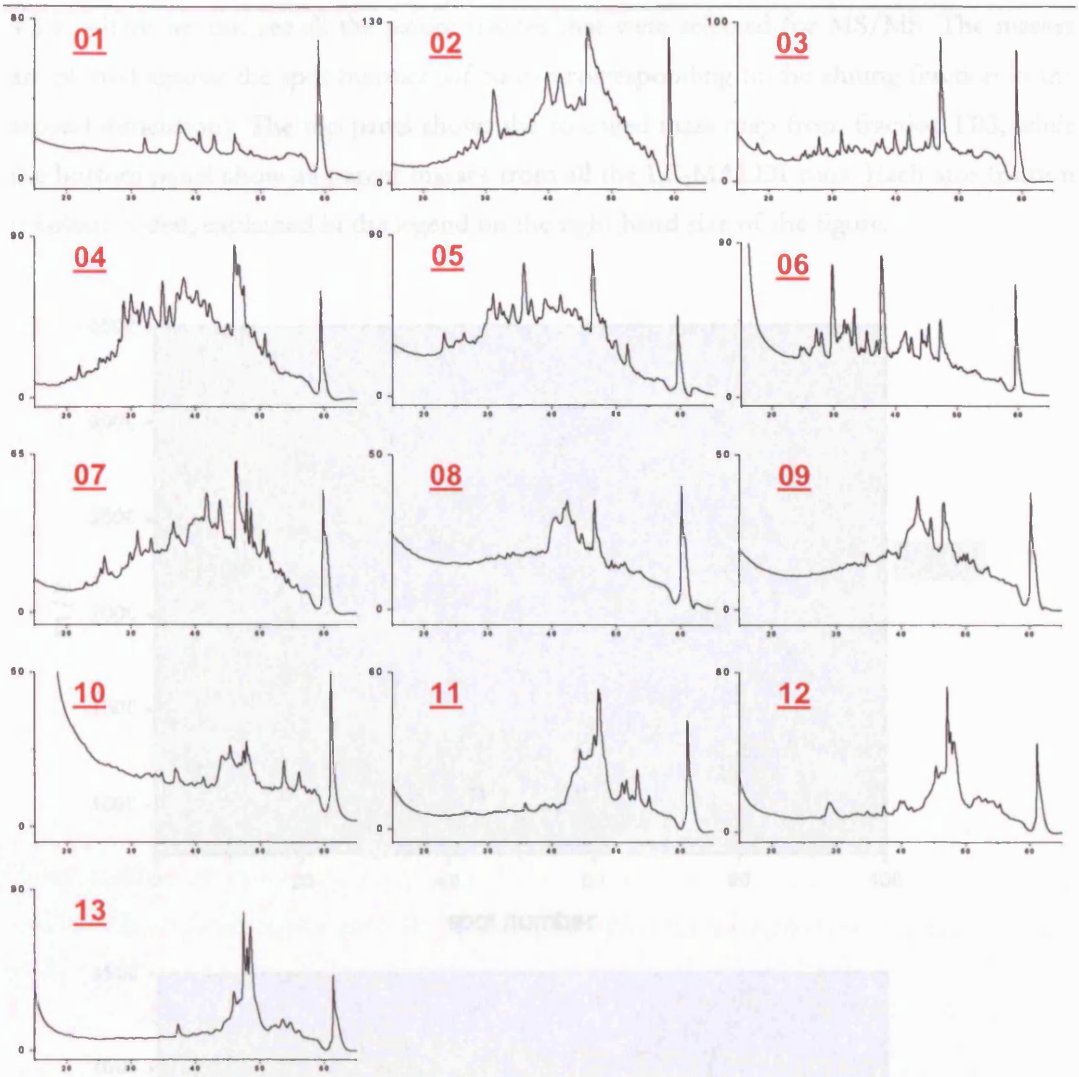


Figure 4.2.1 Second dimension reversed phase separation of *Conus ventricosus* venom.

This figure shows the UV absorbance chromatogram (214 nm) against the spot numbers (effectively time) of the DTT/IAM/trypsin treated fractions from Figure 4.1.1. Concentrated samples were injected onto C18 peptide trap which was offline from the analytical column and washed subsequently with aqueous buffer. The trap was then switched online with the main capillary C18 column and peptides were separated using ACN / H_2O / TFA buffer system. The large peak at the end of each run is that of the organic wash (80 % ACN) partly due to elution polymers, but mainly due to the refractive index change of the solvent. The 13 first dimension fractions were spotted individually in the second dimension over 100 spots, directly onto a stainless steel target, corresponding to approximately 30 seconds fractions. TOF-MS spectra were taken from each spot and tandem mass spectra for the 8 most abundant ions in the survey scan. Spectra were calibrated internally with ACTH (18-39 clip).

4.3 Tandem MS analysis of 2D-LC separated Conus ventricosus venom using MALDI-TOF/TOF MS

The mass spectrometric data obtained from the LC-MALDI approach was overwhelming. Over 5,500 tandem mass spectra were recorded. Some of this data is captured in Figure

4.3.1, where we can see all the parent masses that were selected for MS/MS. The masses are plotted against the spot number (of course corresponding to the eluting fraction in the second dimension). The top panel shows the so called mass map from fraction F03, while the bottom panel show all parent masses from all the LC-MALDI runs. Each size fraction is colour coded, explained in the legend on the right hand size of the figure.

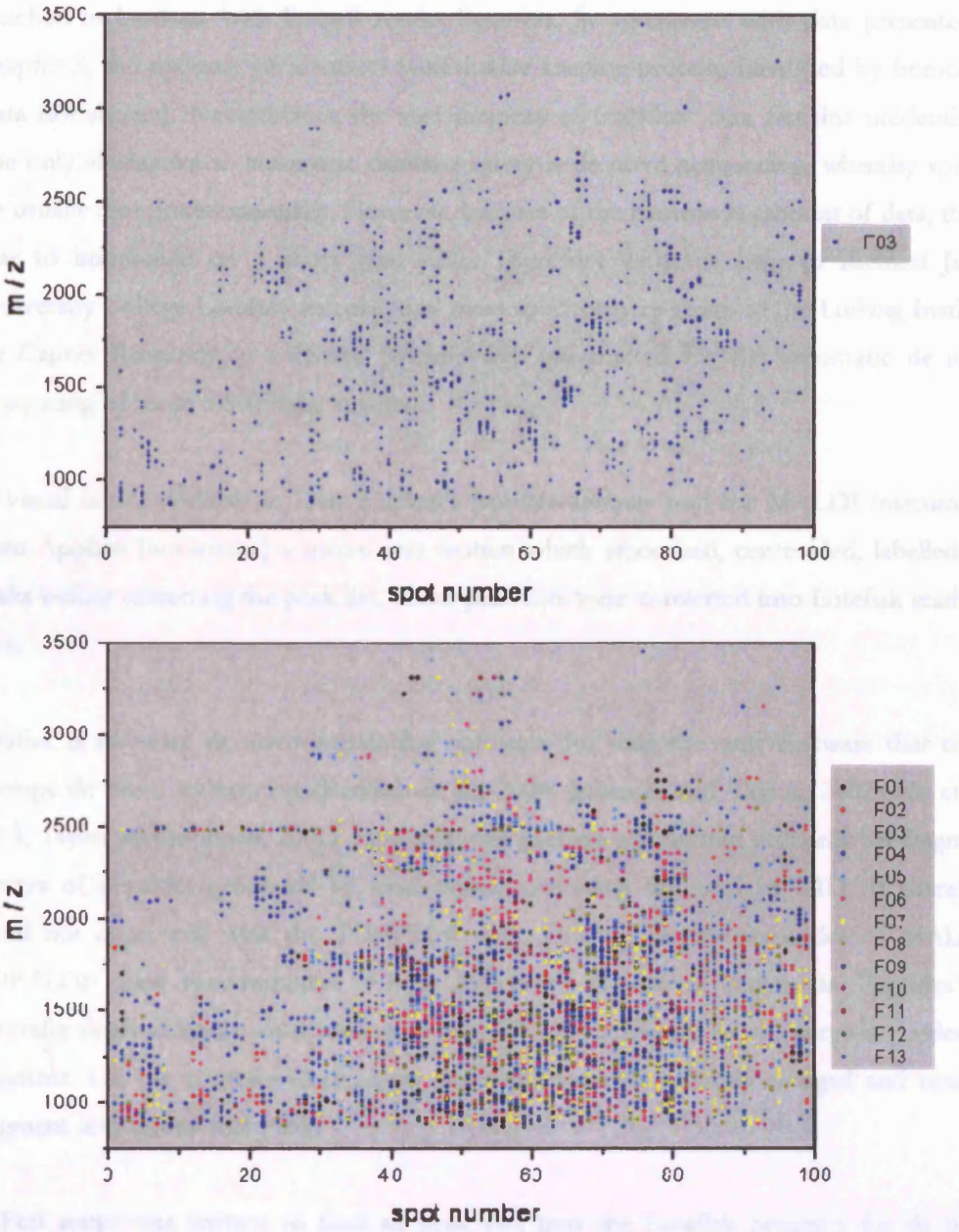


Figure 4.3.1 Mass maps generated by LC-MALDI analysis of *Conus ventricosus* venom.

Peak lists were generated for the LC-MALDI runs of the size fractions from *Conus ventricosus* venom. Each dot in the figures represents the parent ion mass selected for tandem mass spectrometry. In this figure the

mass map from fraction F03 (top) is shown and those of all masses from the 13 fractions combined (bottom). Fractions were spotted every 30 seconds (spot 1 through 99).

The peak lists generated from all the spectra were searched in batches against NCBI using LC-Batch feature of the in-house version of ProteinProspector (Clauser et al., 1999). This software tool allows for the extraction of data from the onboard Oracle database from the ABI 4700. However, the data set was too big to search as a whole. Therefore data was searched in batches, with limited results however. In agreement with data presented in Chapter 3, the majority of identities were house keeping-protein, identified by homology (data not shown). Nevertheless, the vast majority of fragment data remains unidentified. The only alternative to automatic database query is de novo sequencing, whereby spectra are usually interpreted manually. However, because of the enormous amount of data, this is near to impossible on a short time scale. Therefore with the help of Richard Jacob (University College London, Bioanalytical mass spectrometry group in the Ludwig Institute for Cancer Research), a software pipeline was constructed for the automatic de novo sequencing of these 5,500 mass spectra.

In visual basic, available in Data Explorer (spectra analysis tool for MALDI instruments from Applied biosystems) a macro was written which smoothed, centroided, labelled the peaks before extracting the peak list. These peak lists were converted into Lutefisk readable files.

Lutifisk is freeware de novo sequencing software, for long the only software that could attempt de novo sequencing (Banoub et al., 2004; Johnson and Taylor, 2002; Ma et al., 2003; Taylor and Johnson, 2001). However, this package was written primarily for fragment spectra of peptides generated by electrospray ionization followed by CID. It therefore could not cope well with the TOF/TOF data. One of the characteristics of MALDI-TOF/TOF mass spectrometers is high abundance of internal fragments. Peptides are generally singly charged when ionised by MALDI and therefore more energy is needed to fragment. On the contrary electrospray generated ions are multiply charged and tend to fragment into b-ions and y-ion.

A Perl script was written to feed all peak lists into the Lutefisk program for de novo sequencing. An individual text file with the ten best fit sequences was generated for each submitted tandem mass spectrum.

A beta version of Mascot Distiller with de novo sequencing tool was kindly provided by Dr. David Creasy (Matrix Science). As with the Lutefisk program, de novo output was compared with manual annotated spectra. In all the cases, the Matrix Science program performed better. However a major drawback of the Matrix Science program is that it worked stand alone, only allowing to analyse one user loaded spectrum individually, while the Lutefisk program can analyse many spectra in batch mode.

However the data produced by both packages is not shown. The need for user intervention or lack of automatic analysis was considered to be not beneficial for the dataset here at hand.

A second approach to analyse this huge dataset, was to generate a list of top spectra. To this, a Perl script was written (by Richard Jacob) that gave a list of average signal to noise and peak intensity for each tandem mass spectra. With the assumption that a good MS/MS spectrum has a large number of peaks with good S/N spectra were sorted according to average S/N value. However, many of the top 100 'best' spectra contained fragment data from the same peptide.

The analysis of one interesting fraction is shown in Figure 4.3.2, which shows the survey (TOF-MS) spectrum from fraction 3 spot 41. The majority of the peaks visible in this figure are the result of the differential alkylation of the same peptide. This is shown by the inset in the figure, iodoacetamide was used as alkylation reagent, resulting in the 57.0195 Da difference.

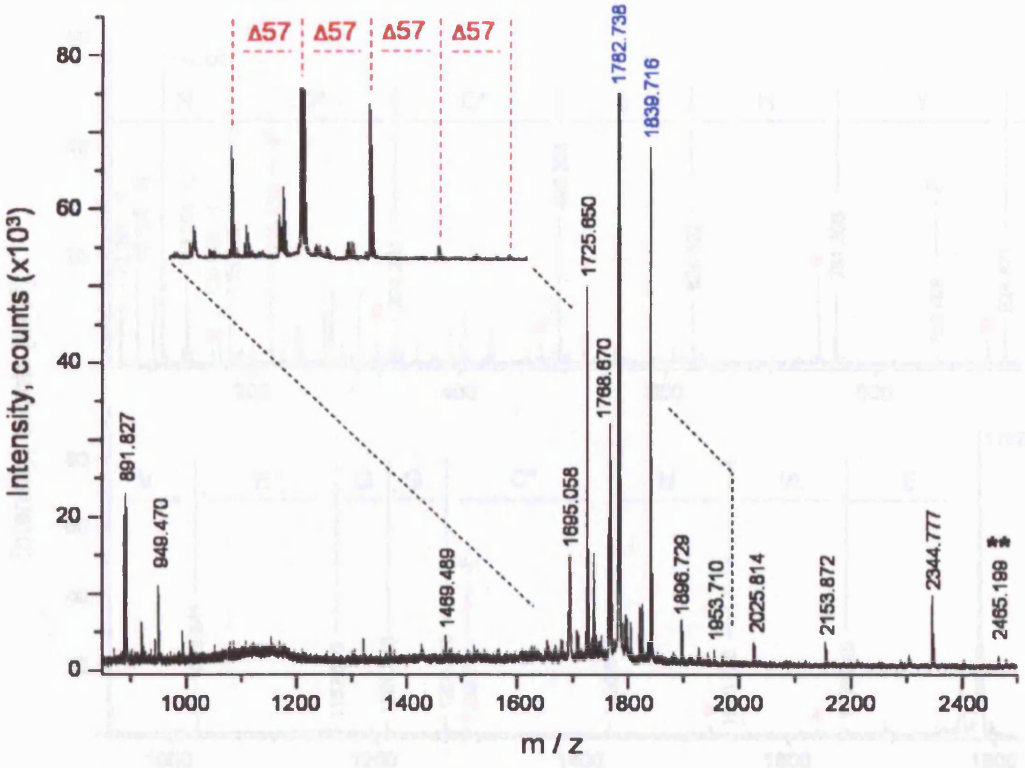


Figure 4.3.2 MS survey scan of fraction F03 spot 41 from *Conus ventricosus*.

This figure shows the TOF-MS spectrum of spot 41 from fraction F03. The peak labelled with (**) is that of the internal standard ACTH. In the insert, a zoom-in of the MS spectrum is displayed. This row of peaks is the result of differential alkylation of this particular peptide. The tandem mass spectra of the two peaks labelled in blue are shown in the next two figures.

From the 8 tandem mass spectra taken from this fraction, two fragment spectra are shown. These are labelled blue in Figure 4.3.2.

Figure 4.3.3 shows the fragment spectrum of m/z 1782.738. The primary sequence was established as:

E(I/L)NC*GGEVYHEC*C*R m/z 1782.700 $[MH]^+$

The measured mass was within 20 ppm of the theoretical mass for this peptide. The alkylated cysteines are displayed as C* while, by mass, no distinction can be made between leucine and iso-leucine (I/L). In the figure immonium ions were labelled in green (the most intense fragment in the entire spectrum was that of the immonium ion of histidine (H)), b and a ions in purple, and y-17 are denoted by a red hash (#). The sequence found, was searched against NCBI using ProteinProspector. However, no hits were obtained.

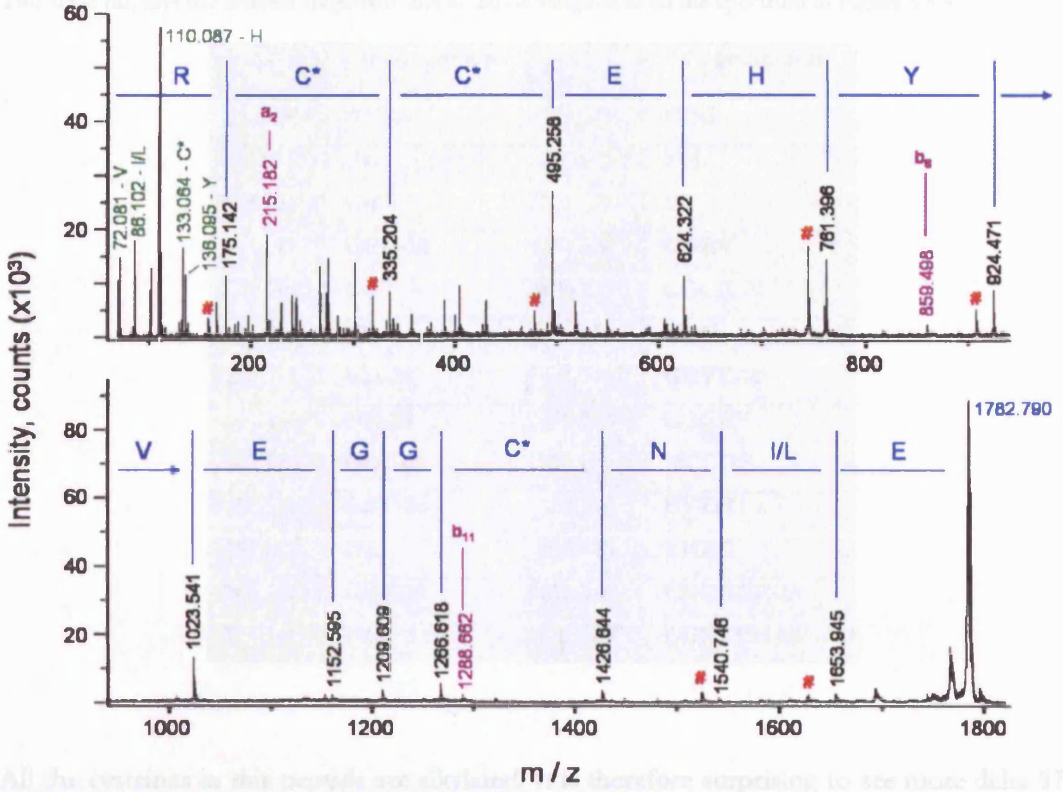


Figure 4.3.3 MS/MS spectrum of m/z 1782.79 from F03-spot 41.

This figure shows the tandem mass spectrum of m/z 1782.790 from F03-spot 41. The spectrum is divided into two sections. The y-ions are labelled in blue. The primary sequence was correctly assigned as E(I/L)NCGGEVYHECCR. Labelled in purple are a-ions and b-ions. Loss of ammonium from y-ions is indicated with a red hash (#). Immonium ions are labelled in green, note the predominant immonium ion of Histidine.

As discussed above, MALDI-TOF/TOF fragmentation generates a high number of internal fragments (Harvey et al., 1995; Medzihradszky et al., 2000). The majority of the peaks at the low end of the mass spectra can be assigned to internal fragment ions. All peaks that were assigned as internal fragments are listed in Table 4.3.1.

Table 4.3.1 Masses of internal fragments.

This table displays the internal fragments that could be assigned from the spectrum in Figure 4.3.3.

m / z	fragment ion	m / z	fragment ion
87.069	GG-28	275.141	CGG
115.051	GG	301.187	YH
187.103	GE	321.129	CC
215.182	GGE-28	343.228	GGEV
218.103	CG	404.214	CGGE
229.165	EV	430.251	YHE
235.172	HE-28	431.240	GEVY-18
247.136	NC-28	503.283	CGGEV
247.136	GGC-28	518.280	NCGGE
258.139	GEV-28	529.320	EVYH
267.157	HE	590.315	YHEC
268.145	GEV-18	613.346	LNCGGE-18
275.141	NC	625.325	GGEVYH-18

All the cysteines in this peptide are alkylated. It is therefore surprising to see more delta 57 adducts. It is know that the N-terminus of peptides can be alkylated upon favourable conditions. However that was not only the case for m/z 1839.716. The tandem mass spectrum of this peptide is shown in Figure 4.3.4. In this figure a mixture of a peptide which was alkylated either on histidine, or at the N-terminal amino group, can be seen.

The y-ion series, as found in Figure 4.3.3, is labelled in blue. However, from the position of histidine, the same sequence can be fitted (with extra alkylation). This is shown is red. Another strong indication of the alkylation of histidine is the immonium ion. Immonium ions of both alkylated and normal histidine were detected, m/z 167.126 and m/z 110.083 respectively. This of course has consequences for database searching, as in general these are not interrogated with variable modifications on histidine.

The b_{13}-ion (m/z 1665.903) confirms that the remaining alkylation (of the peptide with unmodified histidine) can only occur on the N-terminal glutamic acid (E et al.).

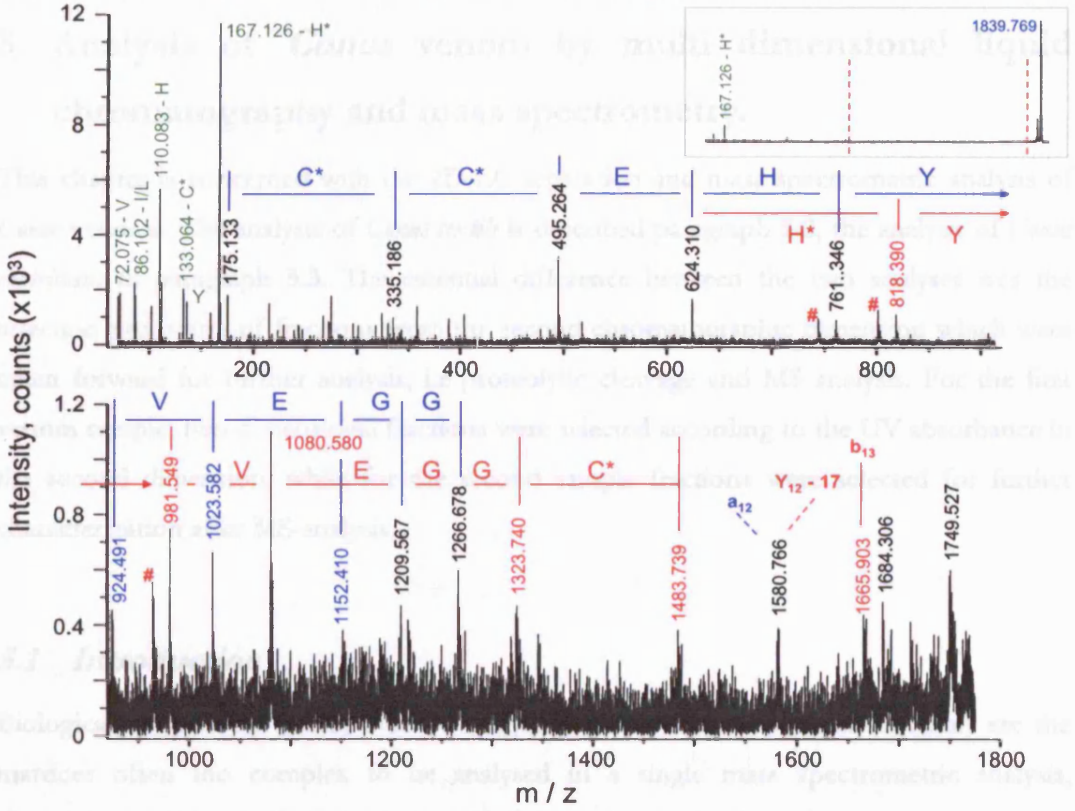


Figure 4.3.4 MS/MS spectrum of m/z 1839.77 from F03-spot 41.

This figure shows the tandem mass spectrum of a mixture of peptides with differential alkylated forms. The sequence series labelled in blue is alkylated on the N-terminal Glutamic acid, while the amino acid sequence labelled in red is alkylated on the Histidine. In both case, the ions series were y-ions. Immonium ions are labelled in green.

The experiments described above were the initial attempt to analyse *Conus* venom using 2D-LC and mass spectrometry. Although the actual identification of putative conotoxins was very low it was a very important experience in sample handling. This data set can be analysed in the future by adequate automatic de novo sequencing program.

5 Analysis of *Conus* venom by multi dimensional liquid chromatography and mass spectrometry.

This chapter is concerned with the 2D-LC separation and mass spectrometric analysis of *Conus* venoms. The analysis of *Conus textile* is described paragraph 5.2, the analysis of *Conus ventricosus* in paragraph 5.3. The essential difference between the two analyses was the selection procedure of fractions from the second chromatographic dimension which were taken forward for further analysis, i.e proteolytic cleavage and MS analysis. For the first venom sample, two-dimensional fractions were selected according to the UV absorbance in the second dimension, while for the second sample fractions were selected for further characterization after MS-analysis.

5.1 Introduction

Biological samples can hardly ever be analysed in their original matrix. Not only are the matrices often too complex to be analysed in a single mass spectrometric analysis, abundant components in the sample will also mask the signals of low abundant species. Furthermore, contaminants such as inorganic salts interfere with ionization processes. For these reasons, fractionation and purification of biological samples are an integral part of mass spectrometric methods.

Comprehensive fractionation also provides a platform for future bioassays to identify biologically active components in *Conus* venom. As previously mentioned, toxins isolated from *Conus* venom target selectively specific ion channels. Because conotoxins can discriminate between closely related subtypes of ion channels, many of the characterised conotoxins are used as pharmacological agents in ion channel research (Moczydlowski et al., 1986) and several have direct diagnostic and therapeutic potential. The targeting specificity derives from the amino acid side chains facilitated by compact globular structure of the conotoxins peptides (McIntosh et al., 1982), which are typically 20 to 30 amino acids long and routinely contain three internal disulfide bridges. The experiments for characterising *Conus* venom have therefore been designed to preserve the conotoxin components in their native form. The sample solubilisation and size separation are performed under high salt concentration to prevent sample aggregation. No reducing agents, such as DTT or TCEP,

are used during the solubilisation and chromatographic separation to keep the disulfide bridges (cystines) intact.

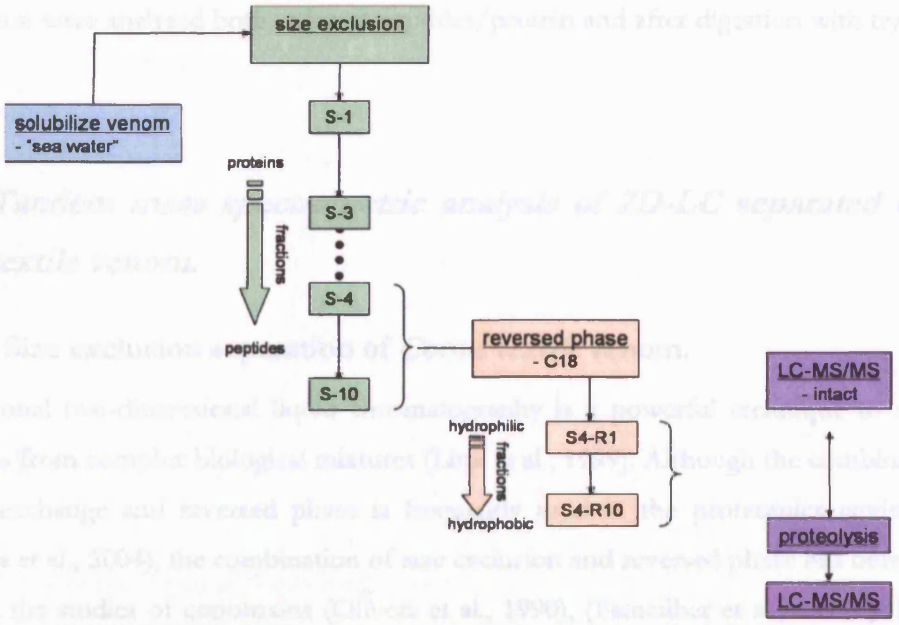


Figure 5.1.1 Workflow for 2D-LC separation of *Conus* venom.

This figure is a workflow of the experiments described in this chapter. Venom was solubilised in artificial sea water (see Figure 2.4.1) and separated (for instance fractions S1 through S10) on a size exclusion column to separate the proteins from the peptides and smaller molecules. Protein fractions were analysed as described in chapter 4. Peptide fractions were desalted using a RP C18 trap column and concentrated under vacuum (for instance fraction R1 through R10 from size fraction 4 (S4)). Each of the size exclusion fractions was separated on a preparative RP C18 column and fractions were taken. All fractions were monitored by MALDI-TOF MS. Fractions from *Conus ventricosus* where first analysed intact by LC-ESI-MS/MS before being treated with a protease.

Figure 5.1.1 Workflow for 2D-LC separation of *Conus* venom.shows a schematic of the workflow adopted for the experiments described in this chapter. Two venom samples were used for analysis, *Conus ventricosus* and *Conus textile* (lyophilised venoms). In brief, the lyophilised venom samples were reconstituted individually in 175 µL artificial sea water (ASW; NaCl, 460 mM; KCl, 10 mM; $CaCl_2$ 11 mM; $MgCl_2$; Tris pH 7.2, 20 mM) and mixed with shaking for 1.5 hours (750 rpm) at 4°C. The sample was spun down and the orange/yellow supernatant was removed. A further 175 µL of ASW was added to the pellet and was left shaking for 1 hour (750 rpm) at 4 °C. This process was repeated one more time. The supernatants were combined to give approximately 520 µL solubilised venom. The solubilised venom was subjected to size exclusion in 5 separate batches of 90 µL each. The retention times of the major peaks of the individual runs were very reproducible. The

first fractions were analysed using 1D-SDS-PAGE as described in chapter 4. The individual peptide fractions from the different elutions were pooled and desalted using a C18 peptide trap. Fractions were concentrated under vacuum to approximately 50 µL and separated further by reversed phase chromatography. Fractions taken from this second dimension separation were analysed both as intact peptides/protein and after digestion with trypsin.

5.2 Tandem mass spectrometric analysis of 2D-LC separated Conus textile venom.

5.2.1 Size exclusion separation of *Conus textile* venom.

Orthogonal two-dimensional liquid chromatography is a powerful technique to separate peptides from complex biological mixtures (Link et al., 1999). Although the combination of cation exchange and reversed phase is frequently used in the proteomics environment (Cutillas et al., 2004), the combination of size exclusion and reversed phase has been longer used in the studies of conotoxins (Olivera et al., 1990), (Fainzilber et al., 1995b). In most cases Sephadex size exclusion resin is used in the first dimension and often only the fractions in the conotoxin mass range that also have biological activity have been analysed. The size exclusion column used in the experiments described in this chapter has an optimal range of 100 – 7,000 Da. In theory when using denaturating conditions (i.e. prevent protein-protein or protein-peptide interaction), proteins co-elute in one fraction while peptides and smaller molecules are separated with higher resolution according to the optimal size separation range of the column.

Figure 5.2.1 shows the absorbance chromatogram at 214 nm of the size separation of *Conus textile*. Fractions that where used for further separation and MS analysis were labelled A07 through B01. Same fractions of consecutive runs were pooled. Individual runs were vary reproducible as shown in Figure 5.2.2. The column was characterised with a known protein mixture as shown in Figure 3.1.1. Therefore components in the molecular weight range 10 kDa to approximately 800 Da were further separated. All known conotoxins have a mass lower then 4,500 Da (Terlau and Olivera, 2004).

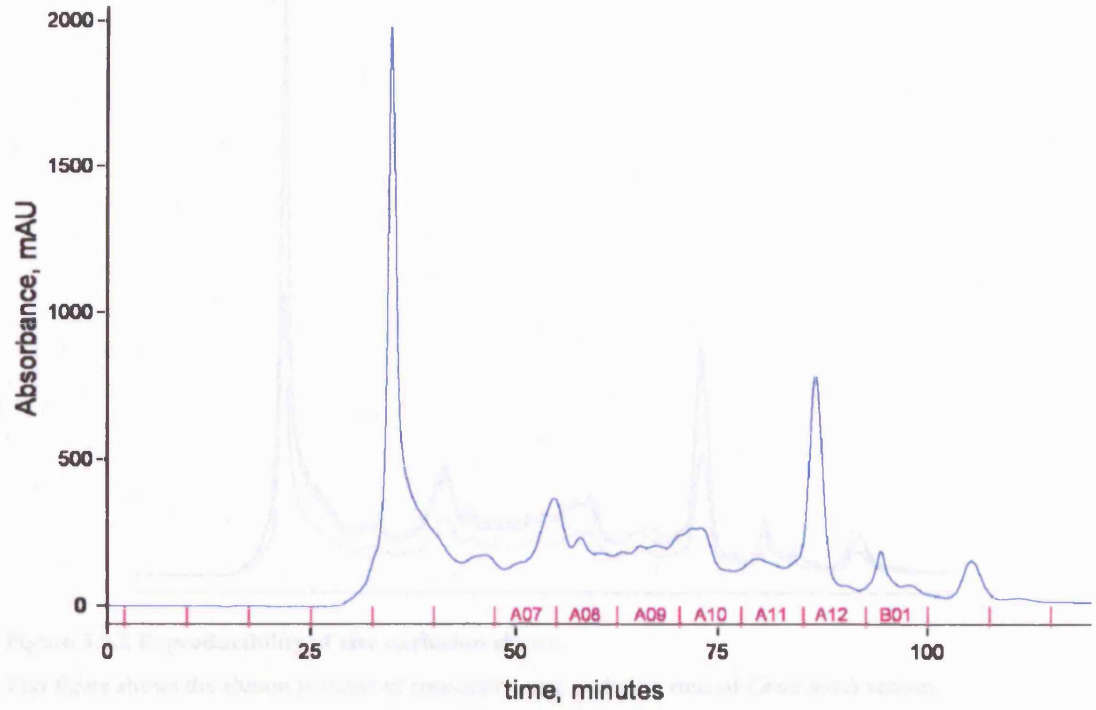


Figure 5.2.1 Size separation of *Conus textile* venom.

The figure shows the absorbance over time for the size separation of *Conus textile* venom. The blue line is the absorbance at 214 nm. Components where separated using an isocratic elution at 0.2 $mL \cdot min^{-1}$ using 150 mM NaCl in 20 mM Tris pH 7.2. Fractions where taken every 1.5 mL (magenta). Fraction A06 through A12 and B01 were further separated by reversed phase chromatography.

Fractions B02, B03 and B04 were cleaned up using the ZipTip protocol and monitored by MALDI-TOF MS, but no components except matrix peaks and chemical noise were detected. The UV absorbance can probably attributed to small organic components and salts from the venom (Escoubas et al., 2000) as well as the Tris from the artificial sea water.

Figure 5.2.3 shows an overlay of the absorbance pattern at 214 nm from Figure 5.2.1 with the online conductivity measurement. This clearly indicates when salt and small conducting organic components eluted. The background conductivity is relatively high at approximately 16.5 $mS \cdot cm^{-1}$. This is caused by the eluent which consists of 150 mM sodium chloride and 20 mM Tris (pH 7.2)

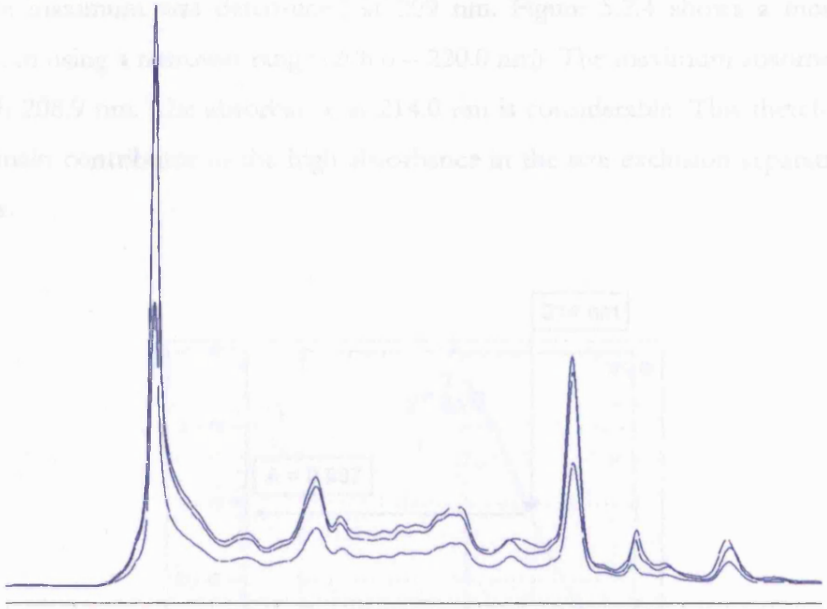


Figure 5.2.2 Reproducibility of size exclusion elution

This figure shows the elution patterns of consecutive size exclusion runs of *Conus textile* venom.

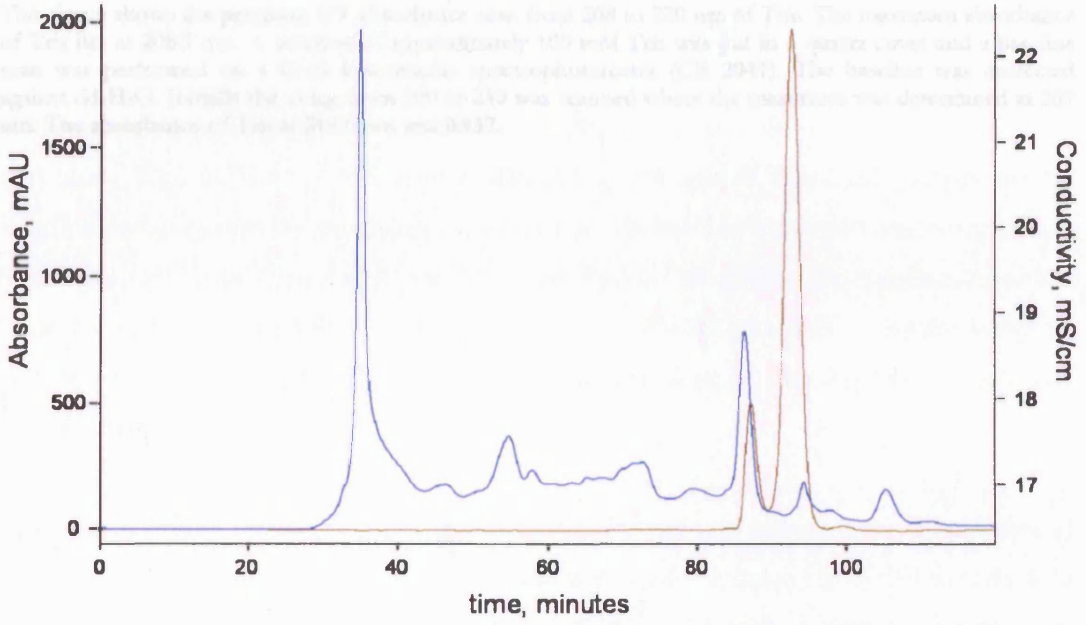


Figure 5.2.3 Conductivity measurements for the size exclusion elution of *Conus* Venom.

This figure shows the absorbance pattern at 214 nm (blue line) of the size exclusion separation of *Conus textile* and the online conductivity measurement (brown). The baseline conductivity is caused by the buffer (150 mM NaCl in 20 mM Tris pH 7.2) used for the isocratic elution. The conductive components are believed to be the salts from the sample and the solubilisation buffer and Tris, which elute roughly in two peaks.

Figure 5.2.4 shows the absorbance profile of Tris (pH 7.2) with varying wavelengths. The absorbance was first scanned on a wider range, namely 200 through 250 nm. Initially the

absorbance maximum was determined at 209 nm. Figure 5.2.4 shows a more accurate zoom-in scan using a narrower range (208.0 – 220.0 nm). The maximum absorbance was at wavelength 208.9 nm. The absorbance at 214.0 nm is considerable. This therefore is most likely the main contributor of the high absorbance in the size exclusion separation around 90 minutes.

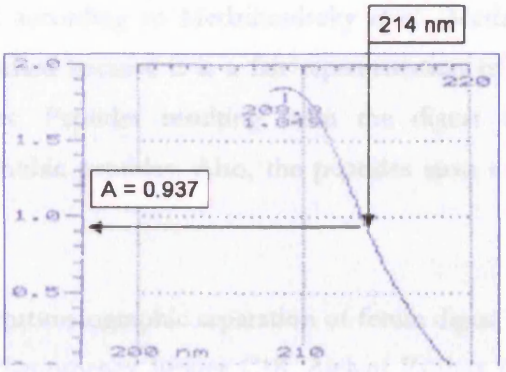


Figure 5.2.4 Zoom in of absorbance spectrum of 100 mM Tris solution.

This figure shows the precision UV absorbance scan from 208 to 220 nm of Tris. The maximum absorbance of Tris lies at 208.9 nm. A solution of approximately 100 mM Tris was put in a quartz cuvet and a baseline scan was performed on a Cecil Instruments spectrophotometer (CE 2041). The baseline was corrected against dd-H_2O. Initially the range from 200 to 250 was scanned where the maximum was determined at 209 nm. The absorbance of Tris at 214.0 nm was 0.937.

5.2.2 RP second dimension separation of size separated *Conus textile* fractions

Before submitting the size fraction described above for further separation by a chromatographic second dimension, the performance of several reversed phase semi preparative analytical columns were evaluated using a trypsin digest of bovine fetuin. Samples were prepared according to Medzihradszky *et al.* (Medzihradszky et al., 1994). Fetuin is used as a standard because it is a fair representation of the elution of peptides from biological samples. Peptides resulting from the digest of fetuin contain both hydrophilic and hydrophobic peptides. Also, the peptides span a wide mass range, from 400 to 7,400 Da.

Figure 5.2.5 shows the chromatographic separation of fetuin digest using three RP columns. From top to bottom, Phenomenex Jupiter C18, Agilent Zorbax C8 and Vydac C18. 250 pmol of fetuin was injected onto the column. The runs of the different columns were performed under the same chromatographic parameter (i.e. flow, gradient, buffer system).

Both the first and last columns are 250 mm in length and have the same internal diameter (4.6 mm). The Zorbax C8 has shorter aliphatic chains and is 150 mm, but the shorter length is compensated for by a smaller particle size. Zorbax has 3 μm particles compared to 5 μm particles from both the Jupiter C18 and Vydac C18. From the three columns the Jupiter had poorest separation performance but retained hydrophilic peptides better in comparison to the Zorbax and Vydac columns. This resulted in a overcrowded middle part of the chromatogram.

The elution profiles of the Zorbax C8 and Vydac C18 are very comparable. The elution of the Vydac C18 seems to be slightly delayed and over a slightly longer time frame, indicating that to some extent this column has a better resolution. The average peak width across the elution was approximately 24 seconds, or 0.2 mL. For the second dimension separation the Vydac C18 column was used and fractions were taken at fixed time intervals corresponding to 600 μl.

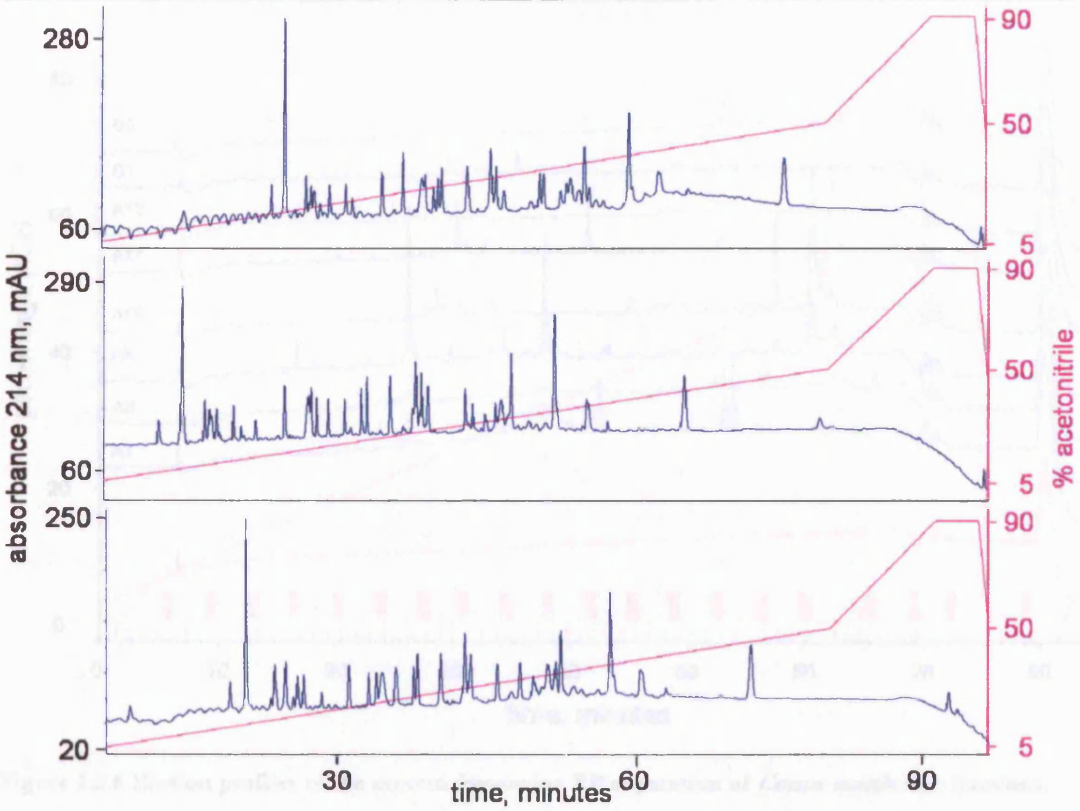


Figure 5.2.5 Evaluation of semi preparative reversed phase columns.

This figure shows the absorbance chromatograms (214 nm) of 250 pmol bovine fetuin digest of different reversed phase columns, namely (from top to bottom) Phenomenex Jupiter C18 (5µm particle size, 4.6 x 250 mm); Agilent Zorbax C8 (3 µm particle size, 4.6 x 150 mm); Vydac C18 (5µm particle size, 4.6 x 250 mm). Peptides were eluted with a flow rate of 0.5mL min^{-1} in a 4 step gradient (magenta lines, 4 minutes 5% B; 39 minutes 50% B; 43 minutes 90% B) using water phase A (0.1% TFA) and organic phase B (0.1% TFA in acetonitrile).

In Figure 5.2.6 the eight second dimension RP elutions A07, A08, A09, A10, A11, A12, B01 and B02 are overlaid, using alternating green and blue colour. As described above, the pooled and concentrated size exclusion fractions (1D) where injected (50 µL). Fractions were taken (magenta numbers on the x-axis) at fixed intervals corresponding to 600 µL into siliconised tubes. So, each size fraction gave rise to 72 reversed phase fractions. As an example, samples were labelled "A09-C11". This then corresponded to first dimension fraction A09 and second dimension fraction C11. In total 576 fractions were taken in the second dimension (8 times 72). A blank, 50 µL of artificial sea water, was run in between each analytical run. One of these runs is shown in magenta, the bottom chromatogram of Figure 5.2.6.

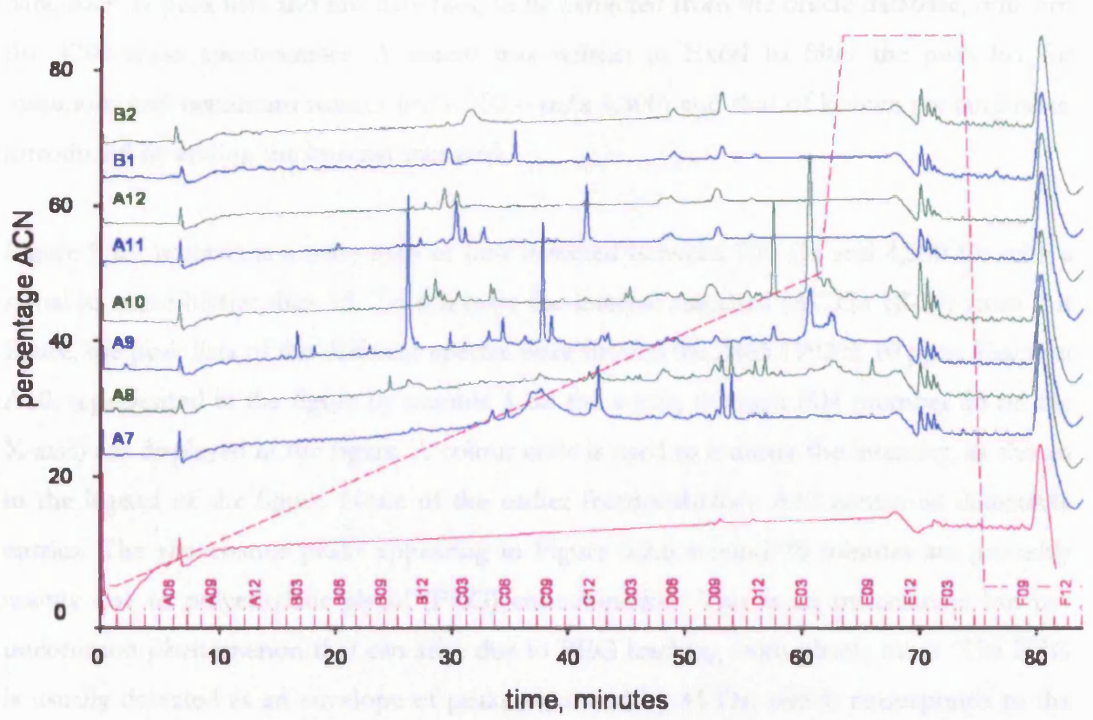


Figure 5.2.6 Elution profiles of the second dimension RP separation of *Conus textile* size fractions.

This figure shows the UV absorbance at 214 nm of the individual chromatograms of the reversed phase separation from size fractions (see Figure 5.2.1) from *Conus textile*. A four step gradient elution was used (dashed magenta line). Fractions were taken every 600 µL as indicated by the magenta indentations on the time axis. Fraction Are named (A01 → A12; B01 → B12 .... F01 → F12). Displayed in solid magenta (the bottom UV trace) is the UV absorbance (214 nm) from a blank run.

In order to get a better idea about how many components there are in *Conus* venom (in the peptide range) and to enhance the detection sensitivity, a mass map was created. All of the approximately 600 fractions were monitored offline by MALDI-TOF MS.

A matrix mixture, consisting of 1 part 70% methanol and 1 part saturated HCCA (Agilent) containing 14 fmol μL^{-1} Adrenocorticotropic hormone clip 18-39 (ACTH), was spotted every 20 seconds by an automated XY-spotter (Probot, LC-Packing / Dionex) on a stainless steel target plate. To 1 µL matrix, 1 µL analyte (2D fraction) was added manually. The analyte/matrix solution was allowed to air dry before analysing with an Applied Biosystems 4700. All fractions of the same size exclusion fraction were loaded on one plate. Spectra were acquired automatically and calibrated automatically using ACTH (m/z 2465.1992). Masses were recorded from m/z 800 to m/z 5,000. Peak lists were extracted using 'Peak Spotter'. Peak Spotter (written by Peter Baker) is a module for the in-house version of ProteinProspector (version 4.11.0) (Clauser et al., 1999). This module allows for

data, such as peak lists and raw data files, to be extracted from the oracle database, onboard the 4700 mass spectrometer. A macro was written in Excel to filter the peak list for minimum and maximum masses (m/z 900 – m/z 4,500) and that of known contaminants, introduced by adding the internal standard.

Figure 5.2.7 represents a mass map of ions detected between 900 Da and 4,500 Da with a signal to noise higher than 15. To eliminate the internal standard (ACTH 18-39) from this figure, the peak lists of the different spectra were filtered for 2465.1992 ± 10 ppm. Fraction A10, represented in the figure by number 1 on the x-axis, through F04 (number 55 on the X-axis) are displayed in the figure. A colour code is used to indicate the intensity, as shown in the legend of the figure. None of the earlier fractionsbefore A10 contained detectable entries. The absorbance peaks appearing in Figure 5.2.6 around 70 minutes are probably mainly due to polyethylene glycol (PEG) contamination. This is an unfortunate but not uncommon phenomenon that can arise due to PEG leaching from plastic tubes. The PEG is usually detected as an envelope of peaks separated by 44 Da, which corresponds to the repeat monomer unit C_2H_4O.

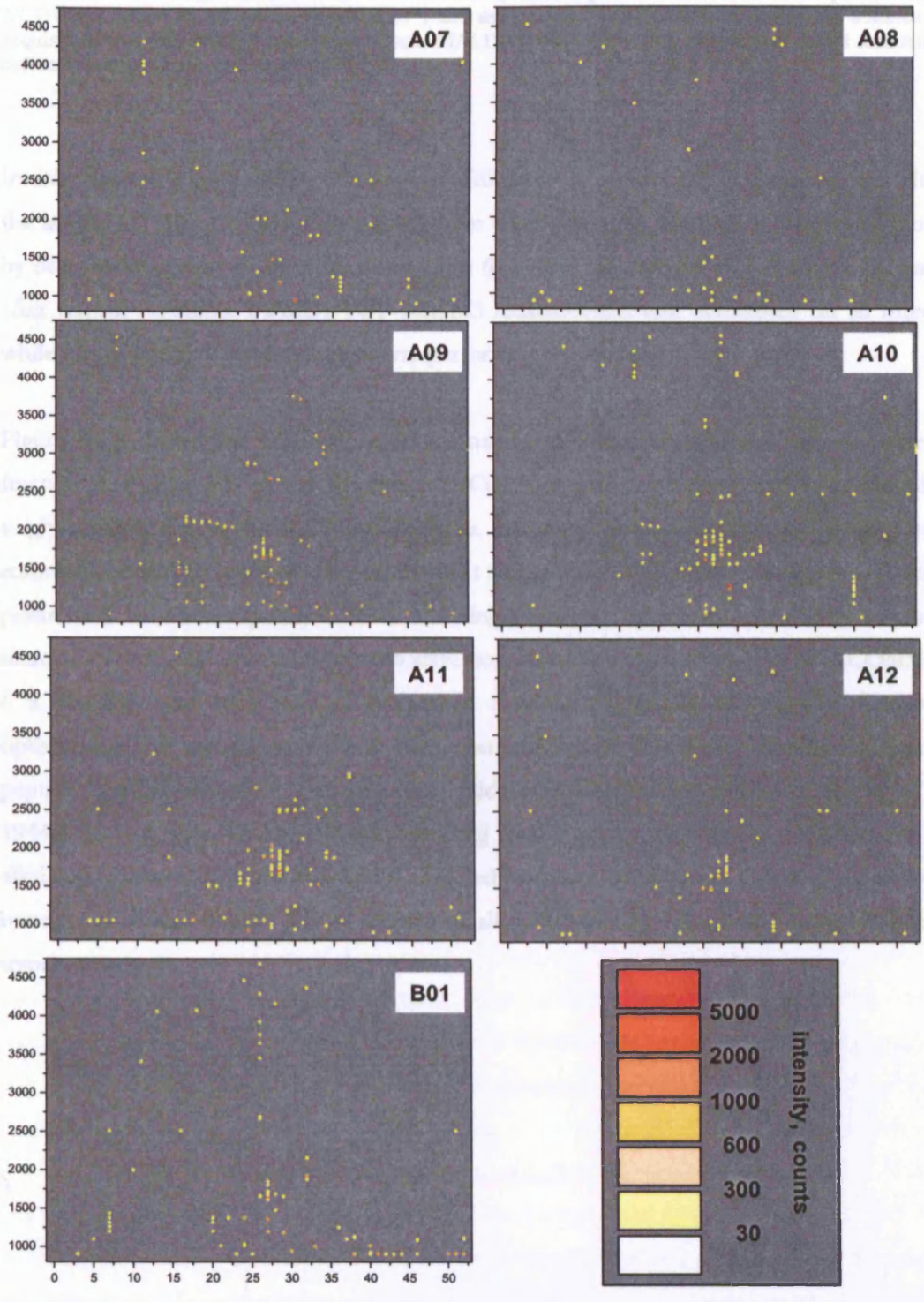


Figure 5.2.7 Mass map of 2D-LC separated *Conustextile* venom.

This figure shows the mass maps of the 2D-LC separation of *Conus textile*. Each individual panel shows the mass to charge ratio measured of each second dimension fraction of *Conus textile*. 1088 components where detected with S/N > 15 for m/z > 900. Intensity is displayed using the colour scheme shown is the

bottom right hand panel of the figure. Matrix was spotted automatically using a Probot™ Micro Fraction Collector (LC Packings - Dionex). Matrix was made up from 1 part 70% methanol and 1 part CHCA. ACTH was added to the matrix (14 fmol·μl^{-1}) and used as an internal standard, during the automatic acquisition with the 4700 Proteomics Analyzer MALDI-TOF/TOF™. 1 µl analyte was added manually before allowing the mixture to air dry.

Initially, from the approximate 600 fractions from 2D LC separation, the ten fractions with the largest UV absorbance (Figure 5.2.6) were taken forwards for further characterization by MS. Samples were analysed in three states (i) native, (ii) reduced and alkylated (iii) and after trypsin digestion. MALDI-TOF MS/MS measurement was performed on all three, while LC-ESI-MS/MS measurement was performed on the latter two.

Figure 5.2.8 shows the TOF-MS screen of native, reduced/alkylated and trypsin treated fraction A11-C02. MS of the fraction A11-C02 identified a predominant monoisotopic singly-charged ion at m/z 1712.4587. After reduction and alkylation and subsequent reanalysis, a shift in mass of the predominant ion to m/z 2060.6414 was observed. Both peaks were mono-isotopically labelled and singly charged. This corresponded to a mass increase of 348.1827 Da indicative of 6 alkylation events with iodoacetamide (CH_2CONH_2; 6 x 58.0293) and suggestive of 6 cysteine residues present in the peptide. Despite optimisation of the reduction and alkylation conditions for highly disulfide knotted peptides, partial alkylation of the parent peptide was also detected as minor ions at m/z 1944.5778 and m/z 2003.6141, corresponding to the ion containing one disulfide and 4 alkylated cysteines, and one free and 5 alkylated cysteines, respectively. Some of the native material remained intact. The reduced and alkylated species remained unaffected after trypsin treatment.

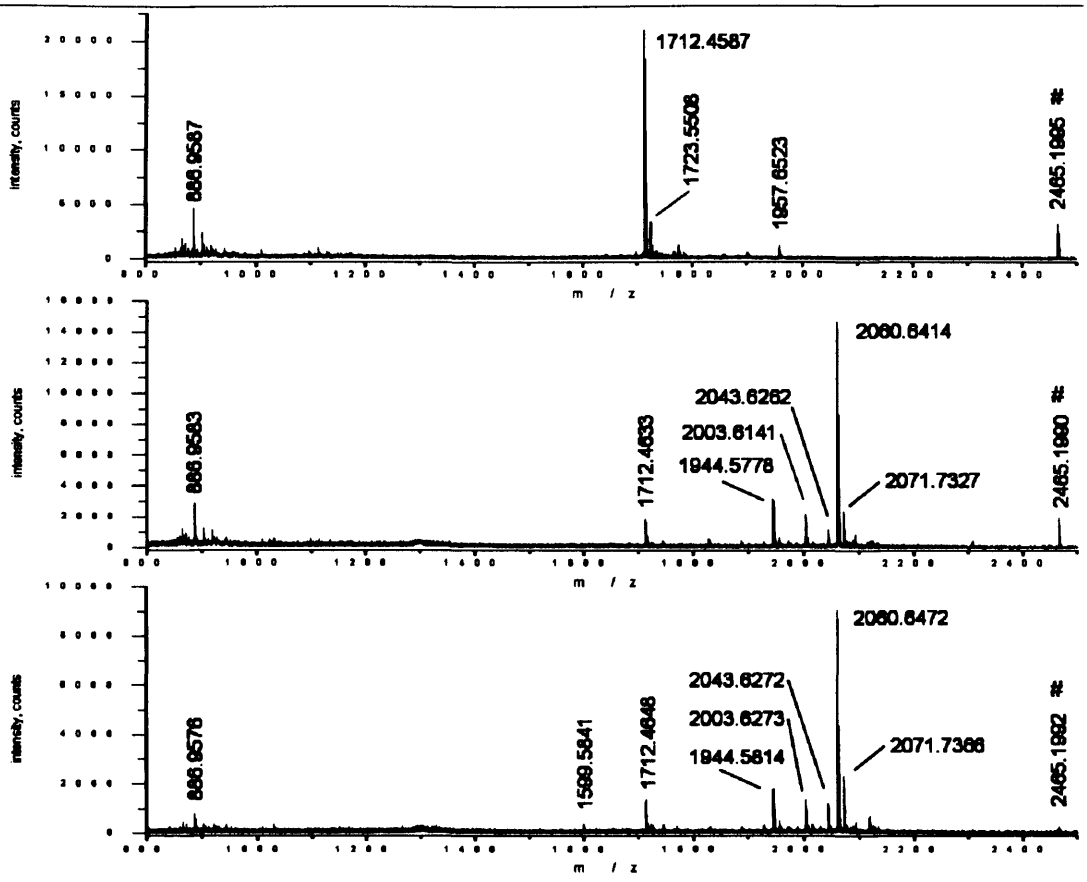


Figure 5.2.8 MALDI-TOF MS spectra of fraction A11-C02 from *Conus textile*

This figure shows the MALDI-TOF MS spectra of fraction A11-C02 from *Conus textile*. The top spectrum is untreated/native sample. The middle panel is reduced (TCEP) and alkylated (IAM) fraction. The bottom panel is after trypsin in solution digestion. The spectrum is calibrated internally on ACTH 18-39 (m/z=2465.1992) which is marked by the blue hash (#). The major ion in the top panel (m/z=1712.4587) shifts 348.1827, indicating 6 alkylations (6 x 58.0293).

In order to elucidate the primary structure, sequence information was collected both from the MALDI and ESI experiments. Tandem mass spectra were collected for both the native and alkylated peptide (see Figure 5.2.9). Fragmentation by MALDI-TOF/TOF of the disulfide linked peptide (top panel) yielded little to no information. Two weak ions where detected which, by mass, can be attributed to the internal fragments of HP and HPS. These internal fragments probably originated from the partially unfolded peptide. This will be discussed later in this paragraph. When fragmenting the alkylated peptide under the same conditions, ions from the b-ion and y-ion series were generated, which can be seen in the bottom panel of the figure.

Protonation of ions occurs mainly at the more basic sites. In the case of peptides, the N-terminal amino group is basic. During positive mode ionization, the additional protons reside on the basic amino acids (R, K, H, P) first and then on the peptide backbone amide groups, which are more statistically distributed along the chain. The presence and position of a basic amino acid in the peptide sequence influences the fragmentation process (Hoffmann and Stroobant). The fragmentation of this singly charged peptide favoured fragmentation at the proline and histidine amino acids that were present in this sequence. It has been reported that the increase of collision energy during low energy CID only decreases the absolute signal intensities of the precursor ion and the high mass fragment ions (Cramer and Corless, 2001).

Proline favours fragmentation at the N-terminal side of the amino acid. This was useful to assign the different ion series. Because of this, the b ion which should have been generated usually lacked or had a significantly lower intensity. In case of the y-ion series, the y-ions before the proline position are completely missing. Although this spectrum gave information about some of the components of the peptide, it was not enough to assign a complete sequence.

The fragmentation pattern of peptides depends not only on the type of amino acid and its location, but also on the experimental fragmentation technique used. In contrast to MALDI ionization, electrospray ionization of peptides (up to 25 amino acids) generally produces doubly charged ions.

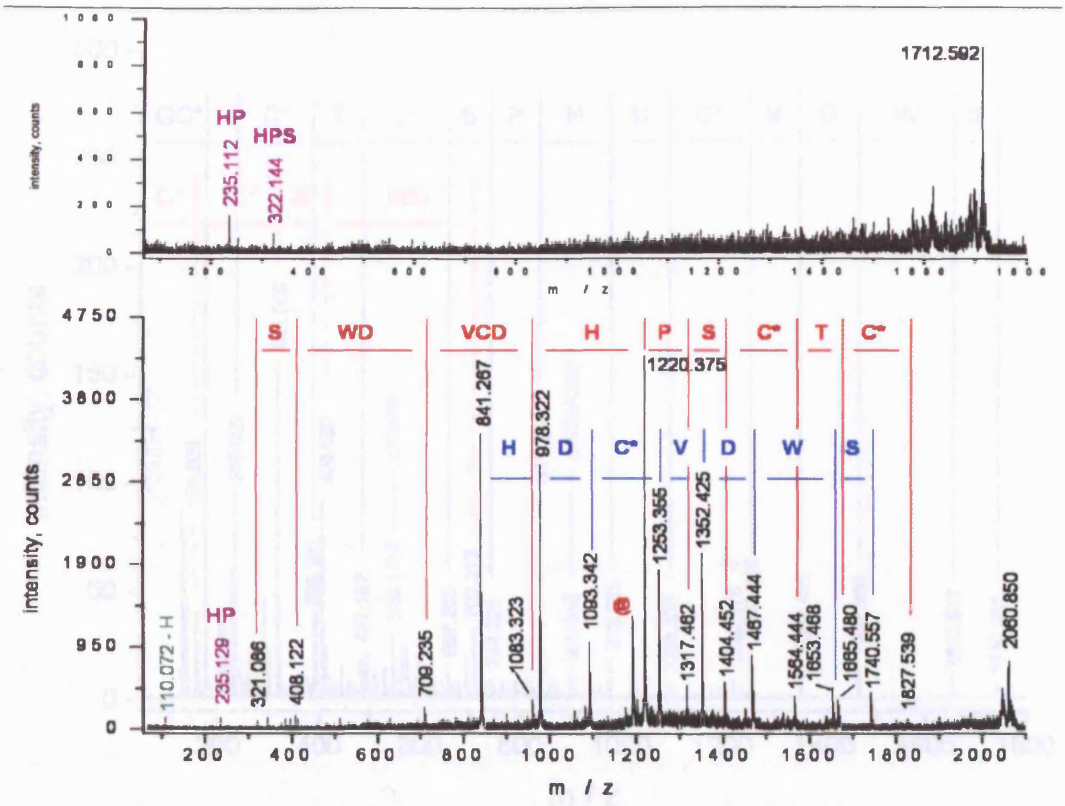


Figure 5.2.9 MALDI tandem MS spectra from fraction A11-C02.

The top panel shows the tandem mass spectrum of m/z 1712.592, from fraction A11-C02 from *Conus textile*. This ion is still disulfide linked. There are two fragment ions which can be assigned as internal fragment ions. The bottom panel shows the MS/MS spectrum of m/z 2060.850 (reduced and alkylated peptide from the top panel). This peptide has also been incubated with trypsin, but no cleavage products resulted, indicating the lack of arginine and lysine amino acids or the presence of RP or KP. The y-ion series are labelled in blue, the b-ion series in red. The immonium ion of histidine is labelled in green and internal fragment ions in purple. Loss of H_2O is indicated by (@).

Typically the two protons allow for a more even fragmentation along the whole peptide backbone. Of course this process is still influenced by sequence specific amino acids. The advantages of using a complementary fragmentation technique is demonstrated in Figure 5.2.10 where the same peptide was dissociated in a CID cell using a quadrupole orthogonal acceleration time of flight (Q-TOF) mass spectrometer.

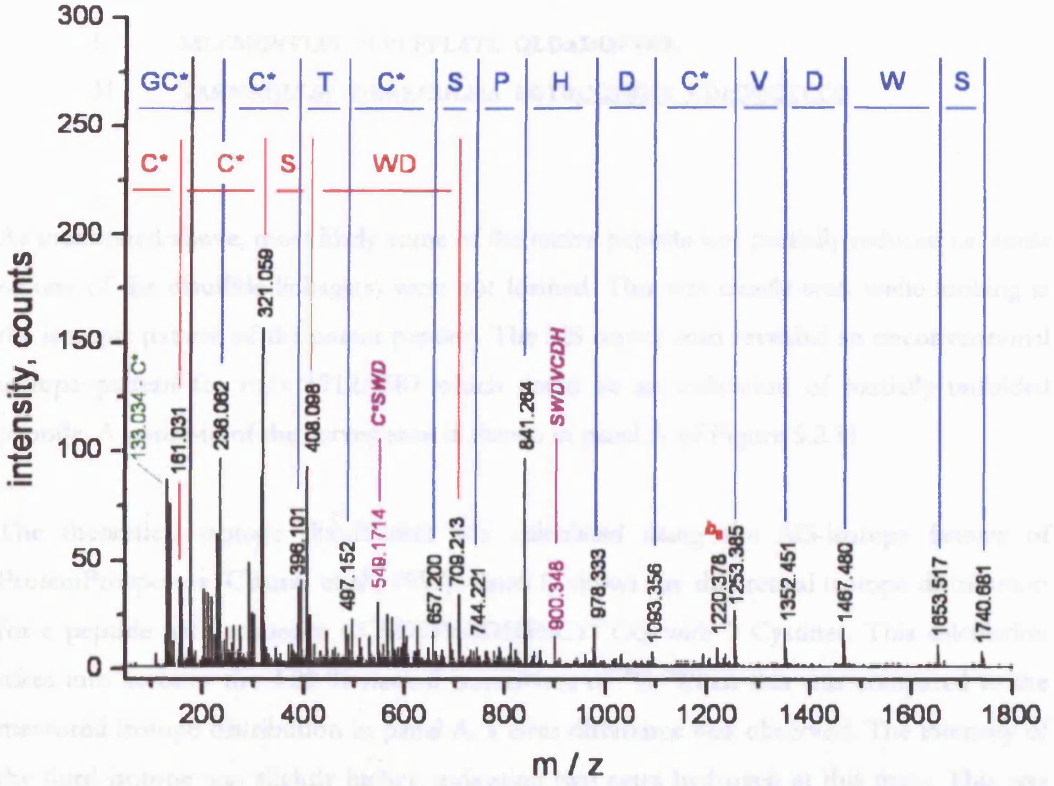


Figure 5.2.10 ESI Tandem mass spectrum of m/z 1030.828 of fraction A11-C02 from *Conus textile* venom.

This figure shows the MS/MS spectrum of m/z 1030.828 $[M+2H]^{2+}$ selected for CID from fraction A11-C02 from *Conus textile*. The y-ion series is labelled in blue, while the b-ions are in red. Internal fragments are labelled in purple and the cysteine immonium ion in green. The C* indicates an alkylated cysteine residue. From this spectrum a complete sequence was assigned (CCSWDVCDHPSCTCCG). No ions were detectable above the noise lever over m/z 1745.

From this tandem mass spectrum the complete primary sequence was assigned, namely: **CCSWDVCDHPSCTCCG.** In comparison to the singly charged MALDI fragment spectrum, low mass fragments ions were generated. In this case the extra proton was retained by the basic amide groups on the protected cysteine residues (C*), introduced by the alkylating reagent. Also internal fragments that contained C* were detected, these were labeled purple in the figure and added to the confidence in assigning this sequence.

The sequence was searched against NCBI using ProteinProspector MS-TAG. (Clauser et al., 1999). The sequence found was the mature toxin part of a conotoxin precursor which was classified in the III/IV scaffold (accession number: AAG60381 - GI number 12619439). The entire precursor predicted by Conticello *et al.* (Conticello et al., 2001) is shown below (70 amino acids).

1 **MLKMGVVLFI FLVLFPLATL QLDADQPVER**

31 **YAENKQLLSP DERREIILHA LGTRCCSWDV CDHPSCTCCG**

As mentioned above, most likely some of the native peptide was partially reduced i.e. some or one of the disulfide linkage(s) were not formed. This was clearly seen while looking at the isotopic pattern of the parent peptide. The MS survey scan revealed an unconventional isotope pattern for m/z 1712.4587 which could be an indication of partially unfolded peptide. A zoom-in of the survey scan is shown in panel A of Figure 5.2.11.

The theoretical isotope distribution was calculated using the MS-isotope feature of ProteinProspector (Clauser et al., 1999). Panel C shows the theoretical isotope distribution for a peptide with sequence CCSWDVCDHPSCTCCG with 3 Cystines. This calculation takes into account the 4.29 % natural occurrence of ^{34}S. When this was compared to the measured isotope distribution in panel A, a clear difference was observed. The intensity of the third isotope was slightly higher, indicating two extra hydrogen at this mass. This was never observed previously for conotoxin peptides.

In order to make a rough estimation of the percentage of unfolded conotoxins, theoretical isotope patterns of mixtures of partially reduced and native conotoxin were calculated. Panel B and D shows the calculated isotope distribution of mixtures of the peptide, with both 2 and 3 Cystines. It was estimated that approximately 6 % of the peptide was in the partial unfolded state with the hypothesis that all forms have the same ionization efficiency. Panel D is 20 % unfolded peptide.

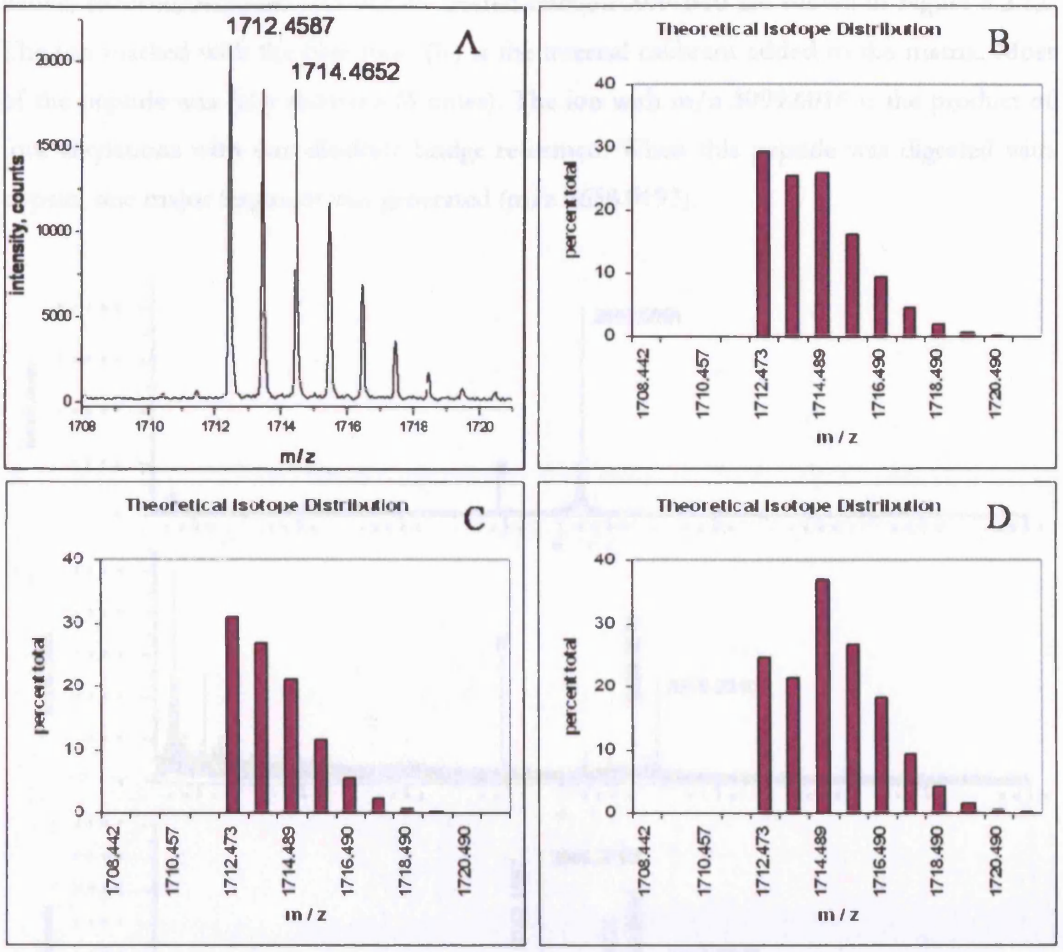


Figure 5.2.11 Theoretical isotope distribution of m/z 1712.458 from A11-C02.

This figure shows the measured and theoretical isotope distribution of m/z 1712.458 from fraction A11-C02 from *Conus textile*. Panel **A** shows a zoom in of the top spectrum of Figure 5.2.8. The isotope pattern shows a distorted pattern, indicating reduced disulfide bridges. Panel **C** shows the theoretical isotope distribution of triply intra disulfide linked peptide $[MH]^+$ CCSWDVCDHPSCTCCG (elemental composition C_{65} H_{90} N_{19} O_{24} S_6). Panel **B** shows a mixture of native peptide, with native peptide with 1 disulfide bond reduced (elemental composition C_{65} H_{92} N_{19} O_{24} S_6). This is a theoretical mixture with 94 % native peptide. Panel **D** shows, as an example a mixture of 80 % native peptide with 20 % 1 reduced disulfide bond. Theoretical distributions were calculated using the MS-isotope feature of ProteinProspector (http://prospector.ucsf.edu/ucsfhtml4.0/msiso.htm).

However, not all elucidations were that straight forward. The MALDI-TOF-MS spectra of native, reduced/alkylated and trypsin treated fraction A09-B10 are shown in Figure 5.2.12. The ion marked with the blue hash (#) is the internal calibrant added to the matrix. Most of the peptide was fully alkylated (6 times). The ion with m/z 3099.6016 is the product of four alkylations with one disulfide bridge reformed. When this peptide was digested with trypsin, one major fragment was generated (m/z 2658.0193).

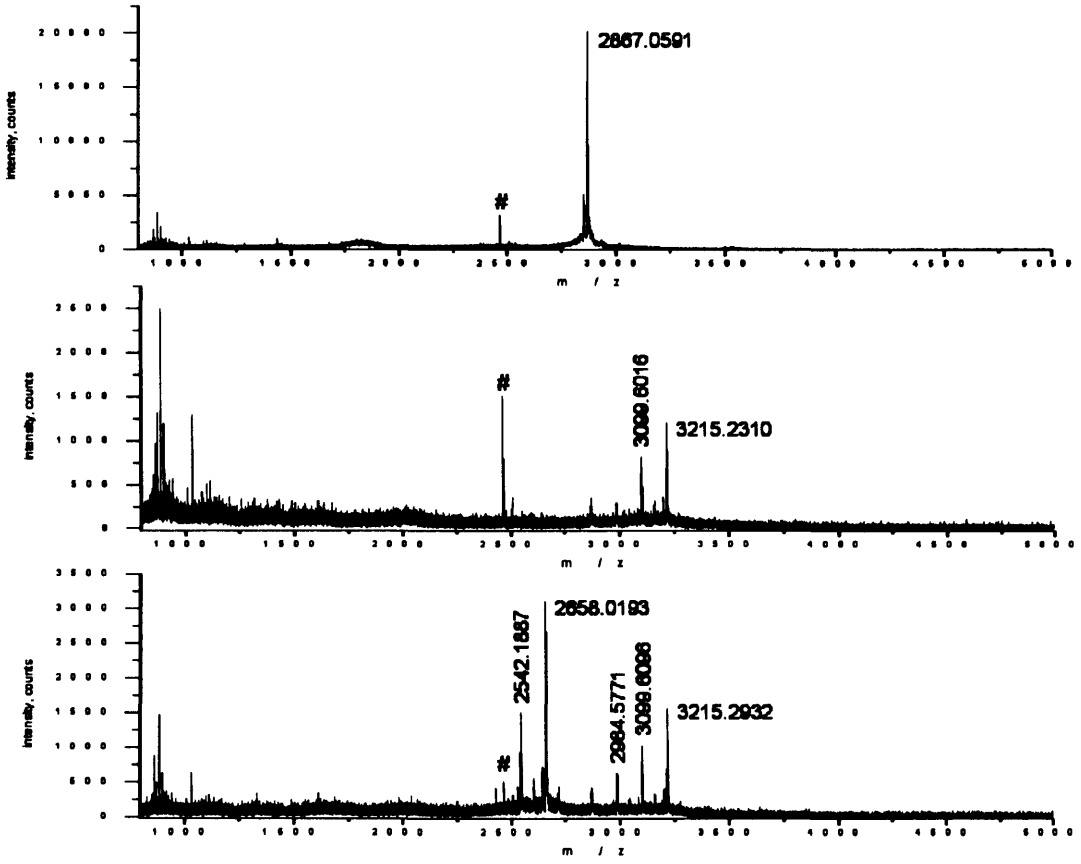


Figure 5.2.12 MALDI-TOF MS spectra of fraction A09-B10 from *Conus textile*

This figure shows the MALDI-TOF MS spectra of fraction A09-B10 from *Conus textile*. The top spectrum is untreated/native sample. The middle panel is reduced (TCEP) and alkylated (IAM) fraction. The bottom panel is after trypsin in solution digestion. The spectrum is calibrated internally on ACTH 18-39 (m/z=2465.1992) which is marked by the blue hash (#). The major ion in the top panel (m/z 2867.0591) shifts 348.1718, indicating 6.000 alkylations (6 x 58.0293).

The tandem mass spectra of m/z 2658.0193 and that of the reduced /alkylated peptide are shown in Figure 5.2.13, top and bottom panel respectively. Both fragment spectra revealed partial but unique information. The preferential cleavage in the sequence is at the C-terminal side of aspartate (D), making the indicated sequence part of the y-ions series.

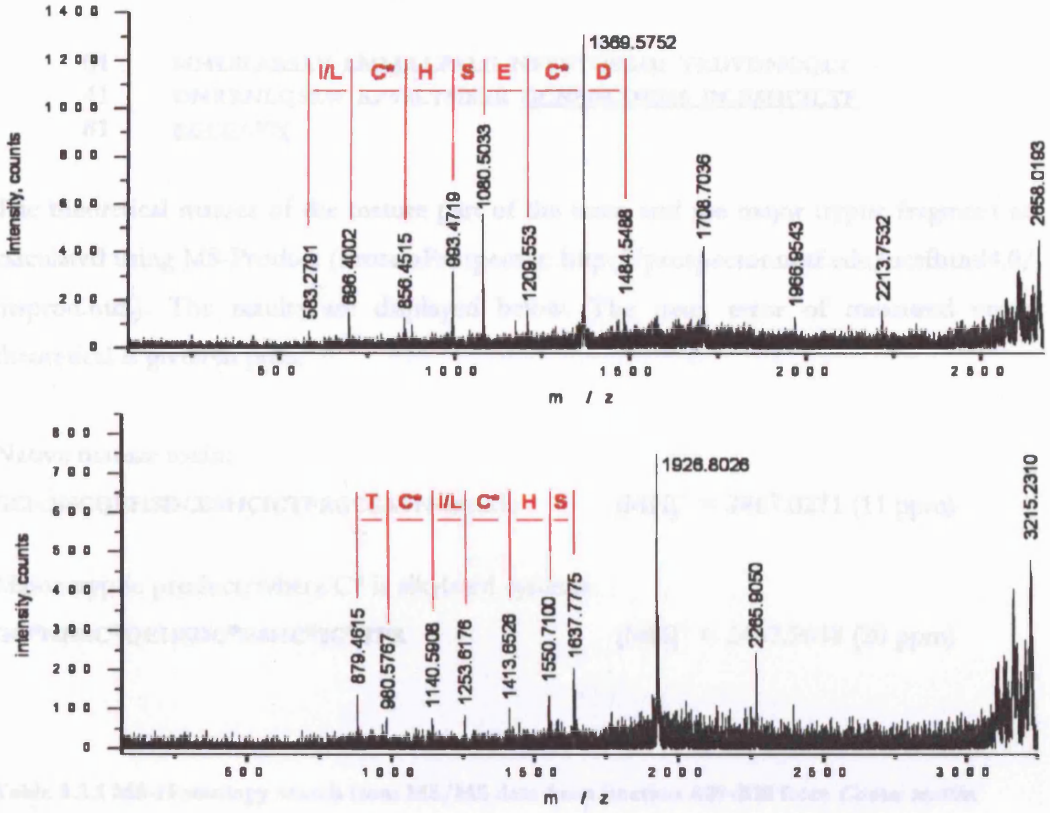


Figure 5.2.13 MALDI tandem mass spectra of peptides from *Conus textile* fraction A09-B10.

Fraction A09-B10 was reduced and alkylated. The tandem mass spectrum of the fully alkylated peptide can be seen in the bottom panel. The MS/MS spectrum of the major proteolytic fragment is shown in the top panel.

The overlapping sequence of both tandem spectra, DCESHCI/LCT, was searched against NCBI using the MS-Homology tool from ProteinProspector (Clauser et al., 1999) (http://prospector.ucsf.edu/ucshtml4.0/mshomology.htm). MS-Homology allows for a user defined number of amino acid substitutions, whereby leucine and isoleucine are represented as one amino acid using the wild card [I|L]. The results of this search can be found in Table 5.2.1. The sequence determined was unique for Spasmodic peptide tx9a precursor (Pi-conotoxin TxIXA). This sequence was first found in *Conus textile* by Lirazan *et al.* (Lirazan et al., 2000). For the search against NCBI two amino acid substitutions were allowed. One further conotoxin met the criteria of the sequence, with one amino acid

substitution. Spasmodic protein tx9a-like protein precursor (Conotoxin gm9a) from *Conus Gloriamaris* (Miles et al., 2002).

The full sequence of pi-conotoxin TxIXA precursor is displayed below. The mature part of the toxin is underlined.

01 **MHLSLARSAV LMLLLLFALG NFVVVQSGQI TRDVDNGQLT**
41 **DNRRNLQSKW KPVSLYMSRR GCNNSCQEHS DCESHCICTF**
81 **RGCGAVN**

The theoretical masses of the mature part of the toxin and the major tryptic fragment are calculated using MS-Product (ProteinProspector: http://prospector.ucsf.edu/ucsfhtml4.0/msprod.htm). The results are displayed below. The mass error of measured versus theoretical is given in ppm.

Native mature toxin:

GCNNSCQEHSDCESHCICTFRGCGAVN-amide $[MH]^+$ = 2867.0271 (11 ppm)

Major tryptic product, where C* is alkylated cysteine:

GC*NNSC*QEHSDC*ESHC*IC*TFR $[MH]^+$ = 2657.9648 (20 ppm)

Table 5.2.1 MS-Homology search from MS/MS data from fraction A09-B10 from *Conus textile*.

The table shows the top four hits for the homology search from the sequence DCESHCIC sequenced from the tandem mass spectra from A09-B10 from *Conus textile*. The homology search was performed with the MS-Homology package from ProteinProspector. For the search two amino acid substitutions were allowed. Amino acid substitutions are displayed in bold. The sequence was unique for a precursor conotoxin π-TxIXA from *Conus textile*.

Matching Sequence	Species	Protein Name
(S)DCESHCIC(T)	*Conus textile*	Spasmodic peptide tx9a precursor (Pi-conotoxin TxIXA)
(S)DCASHCIC(T)	*Conus* Gloriamaris	Spasmodic protein tx9a-like protein precursor (Conotoxin Gm9.1) (Conotoxin gm9a)
(A)DCRSNCIC(E et al.)	Desulfitobacterium hafniense DCB-2	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 1
(T)DCNSHCHC(E et al.)	Anopheles Gambiae Str. Pest	ENSANGP00000011827

The tandem ESI mass spectrum of the tryptic fragment is shown in the top panel of Figure 5.2.14. Although the ion has a substantial response in MS mode (bottom panel), the fragmentation spectrum is poor. The major fragment ion in the spectrum can be contributed to the b_2 ion (C*G). No fragmentation spectra were recorded for either the intact alkylated mature toxin, or lower charged tryptic fragments of this peptide. Fragmentation of peptides during mass spectrometry, and the energy required for fragmentation, has some dependency on the peptide sequence, but also has a correlation with the peptide mass, where larger peptides require more collision energy.

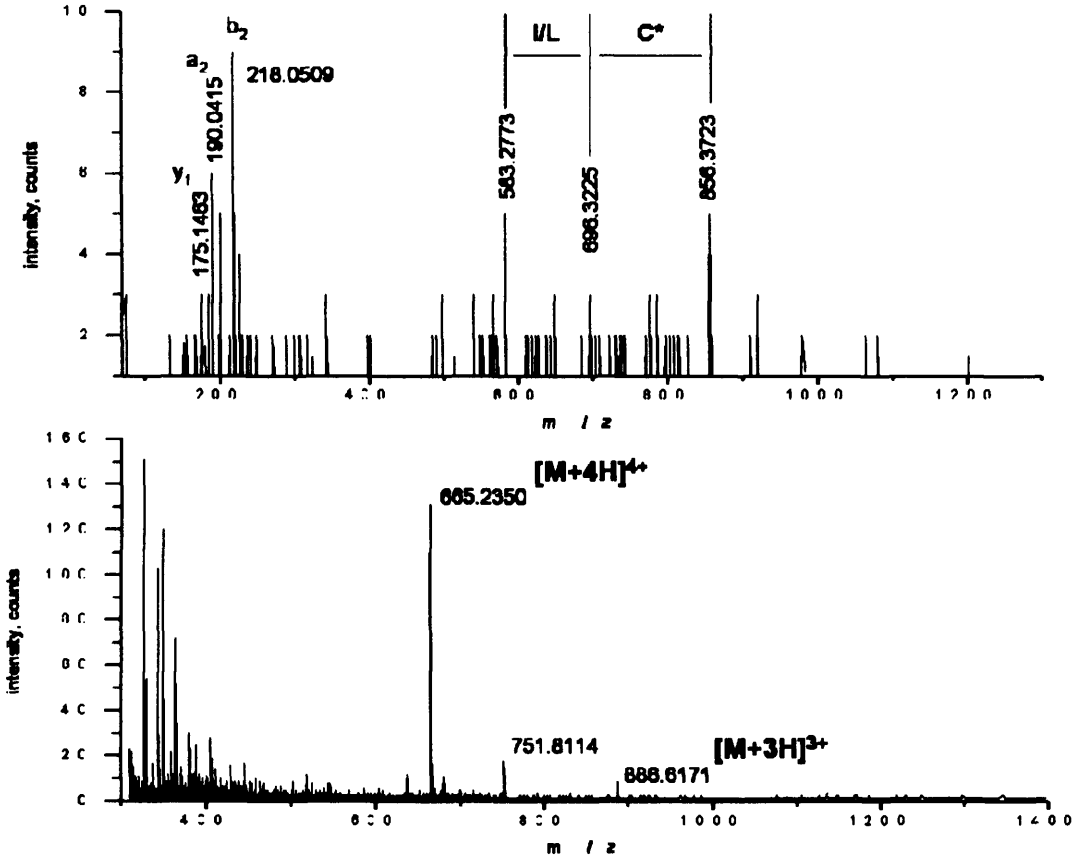


Figure 5.2.14 Tandem ESI mass spectrum m/z 665.2350 of *Conus textile* fraction A09-B10.

This figure shows the MS scan (bottom) and MS/MS spectrum (top) of the major tryptic fragment from fraction A09-B10 (gc*nnsc*qehsdc*eshc*ic*tfr [MH]+ = 2658.0193). The y-ions are displayed in blue and the b-ions in red. Although there is a substantial signal in MS mode, the fragment spectrum is rather poor. However the fragments were consistent with those of the MALDI-MS/MS spectrum of the same peptide. No fragmentation was observed for the triply charged ion or of the intact alkylated peptide before trypsin digestion

To validate the assignment of pi-conotoxin TxIXA found in this fraction, the theoretical fragments that were generated with MS-Product from pi-conotoxin TxIXA of the intact alkylated mature toxin are compared with the experimental fragments. The theoretical fragments are given in Table 5.2.2 Theoretical y-ion series from native mature toxin from A09-B10. The unassigned ions of the MALDI fragment spectrum in the bottom panel of Figure 5.2.13 can now be correlated (m/z 1926.78 and 2265.90) An overall sequence coverage of 35 % was achieved (9 out of 26 y-ion fragments), therefore a confident assignment of the sequence.

Table 5.2.2 Theoretical y-ion series from native mature toxin from A09-B10

To validate the assignment of pi-conotoxin TxIXA to fraction A09-B10, the MALDI tandem mass spectrum is compared with the theoretical y-ion series for this peptide, seen in the bottom pannel of Figure 5.2.13. Nine out of 26 fragment y-ions were detected, yielding 35 % sequence coverage.

Theoretical y-ion series fragment masses $[MH]^+$					
y_1	132.08	y_{10}	**1140.50**	y_{19}	**2265.90**
y_2	231.15	y_{11}	**1253.59**	y_{20}	2394.94
y_3	302.18	y_{12}	**1413.62**	y_{21}	2523.00
y_4	359.20	y_{13}	**1550.68**	y_{22}	2683.03
y_5	519.23	y_{14}	**1637.71**	y_{23}	2770.06
y_6	576.26	y_{15}	1766.75	y_{24}	2884.11
y_7	732.36	y_{16}	**1926.78**	y_{25}	2998.15
y_8	**879.43**	y_{17}	2041.81	y_{26}	3158.18
y_9	**980.47**	y_{18}	2128.84	y_{27}	3215.20

5.3 Comprehensive mass spectrometric analysis of 2D-LC separated Conus ventricosus venom.

The basis of analytical separation was performed in a similar fashion to the separation of *Conus textile* venom in section 5.2. The main difference was that fractions were selected according to mass abundance i.e. number of ions detected in a particular fraction after MALDI-TOF-MS scanning of all fraction As opposed to selection of fractions by UV absorbance. This resulted in manual interpretation LC-MSMS runs of 96 fractions selected.

5.3.1 Size exclusion and reversed phase 2D-LC separation of *Conus ventricosus* venom.

Lyophilised *Conus ventricosus* was solubilised as decribed before and was separated using size exclusion chromatography. The UV absorbance (214 nm) chromatogram is shown in Figure 5.3.1. In comparison with the elution profile of *Conus textile* venom, which was run under identical condition, the chromatograms have great similarities. As was shown in Chapter three, all proteins eluted in fraction A04, A05 and A06. In a similar fashion, a large absorbance peak appeared around 85 minutes. Again, not many components were detected around this area. As was shown in Figure 5.2.3 all salts and most likely all the small organic compounds as well as the solubilisation buffer (ASW) eluted around this time point. The conductivity chromatogram for this run is not shown.

The samples from the same second dimension fractions from five consecutive runs were pooled and concentrated under vacuum. Peptide samples (fraction A06, A07, A08, A09, A10, A11, A12 and B01) were desalted using a C18 peptide trap column. Further separation and analysis of protein fractions in described in chapter 3.

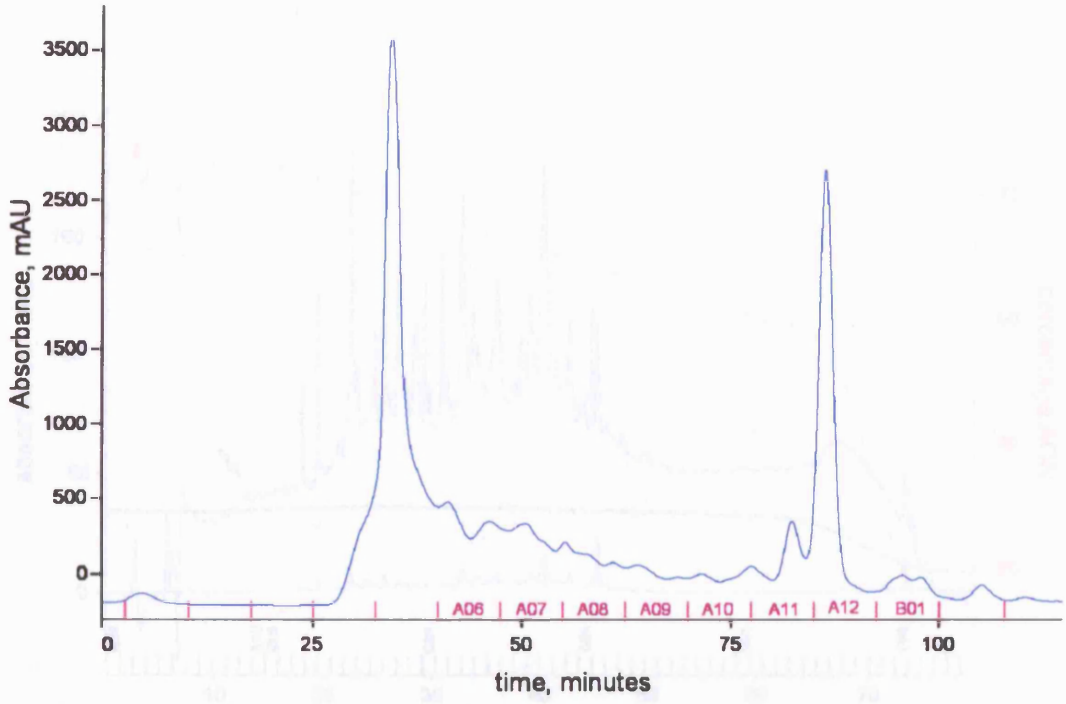


Figure 5.3.1 Size separation of *Conus ventricosus* venom.

The figure shows the UV absorbance over time of the size separation of *Conus* ventricosus venom. The blue line is the absorbance at 214 nm. Components were separated using an isocratic elution at 0.2 mL min^{-1} using 150 mM NaCl in 20 mM Tris pH 7.2. Fractions were taken every 1.5 mL (magenta). Fraction A06 through A12 and B01 were further separated by reversed phase chromatography.

After desalting, the samples were concentrated to approximately 50 µl and injected onto semi preparative C18 column. A detailed chromatogram of one of these fractions (fraction A09) can be seen in Figure 5.3.2. A number of parameters were recorded during the run. The absorbance was recorded with three wavelengths, of which two are displayed in the figure (214 nm light blue and 280 nm dark blue). The pressure (green line) over the column remains constant during the run, until the gradient (dashed magenta line) reaches the high percentage acetonitrile. The conductivity was decreasing slowly with lower percentage aqueous buffer. The fluctuation of the absorbance indicated by the black arrow was apparatus specific. Also in the blank runs, which were run in between the analytical runs, this fluctuation was present. It is believed that this is due to less efficient mixing of the aqueous and organic buffer during the start of the gradient. This can be seen also by the fluctuation of the absorbance during the start of the gradient. When the gradient reached the critical part of the separation, both absorbance and conductivity stabilised. It is not assumed that this influenced the separation quality.

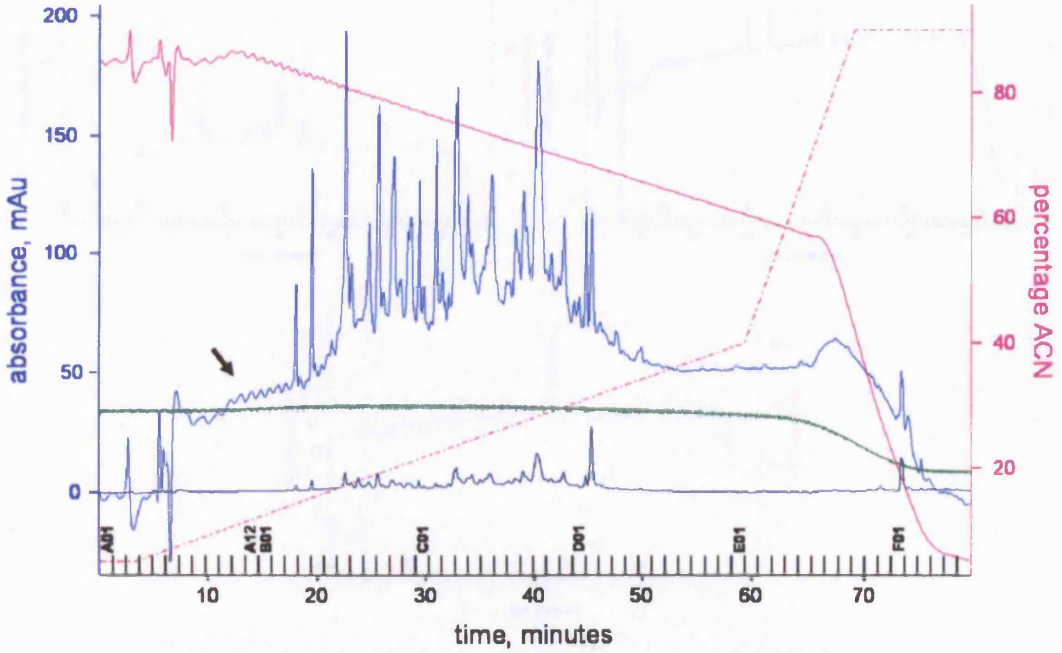


Figure 5.3.2 Reversed phase separation of A09 size fraction from *Conus ventricosus*.

The figure shows the second dimension separation of size fraction A09 (see Figure 5.3.1). The elution profile was recorded at 214 nm (blue) and 280 nm (black). A four step gradient elution was used (dashed magenta line, right Y-axis) Fraction where taken every 600 µl as indicated by the black indentations on the time axis. Fraction Are named (A01 → A12; B01 → B12 …. F01 → F12). The fluctuation seen in the chromatogram at 214 nm (see arrow) is attributed to poor mixing of solvents in the beginning of the gradient. The pressure (green line) remains constant during this part of the separation, but the conductivity also fluctuates (solid magenta line).

The chromatograms of fraction A10, A12 and B01 are shown in Figure 5.3.3. There was a clear difference in the abundance of peaks. Fractions were taken (shown in black numbers on the x-axis in both figures) at fixed intervals corresponding to 600 µL, into siliconised tubes.

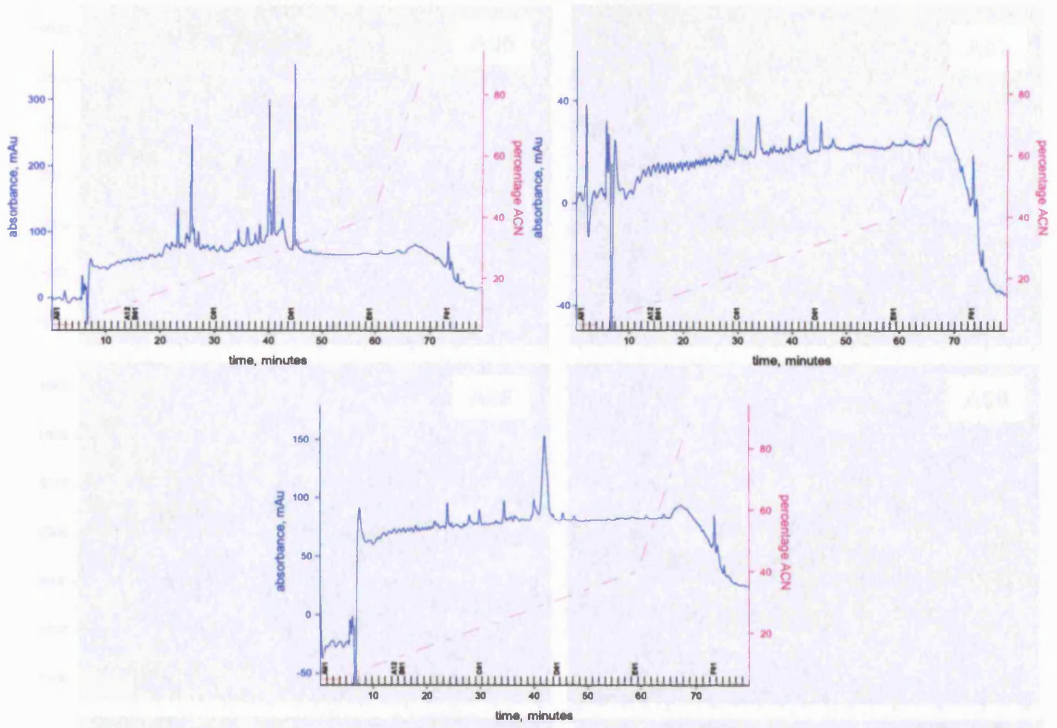


Figure 5.3.3 Reversed phase separation of A10, A12 and B01 size fractions from *Conus ventricosus*.

The figure shows the second dimension separation of size fraction A10, A12 and B01 (clockwise starting from top left). See Figure 5.3.1 for the chromatogram of the size separation. The elution profile was recorded at 214 nm (blue). A four step gradient elution was used (dashed magenta line, right Y-axis) Fractions were taken every 600 µl as indicated by the black indentations on the time axis.

Fractions were labelled as described before. In a similar fashion to the separation of *Conus textile* a mass map was created. 1 µl of each fraction was added manually to 1 µl matrix (spotted automatically) on a stainless steel MALDI target. The matrix consisted of 14 fmol μl^{-1} ACTH 18-39 in one part 70 % methanol and 1 part Agilent CHCA. Intensity was recorded again mass to charge ratio using the Applied Biosystems 4700 proteome explorer.

The mass to charge ratio with intensity is plotted for each second dimension fractions, grouped for each first dimension fraction. The so called mass maps are displayed in Figure 5.3.4.

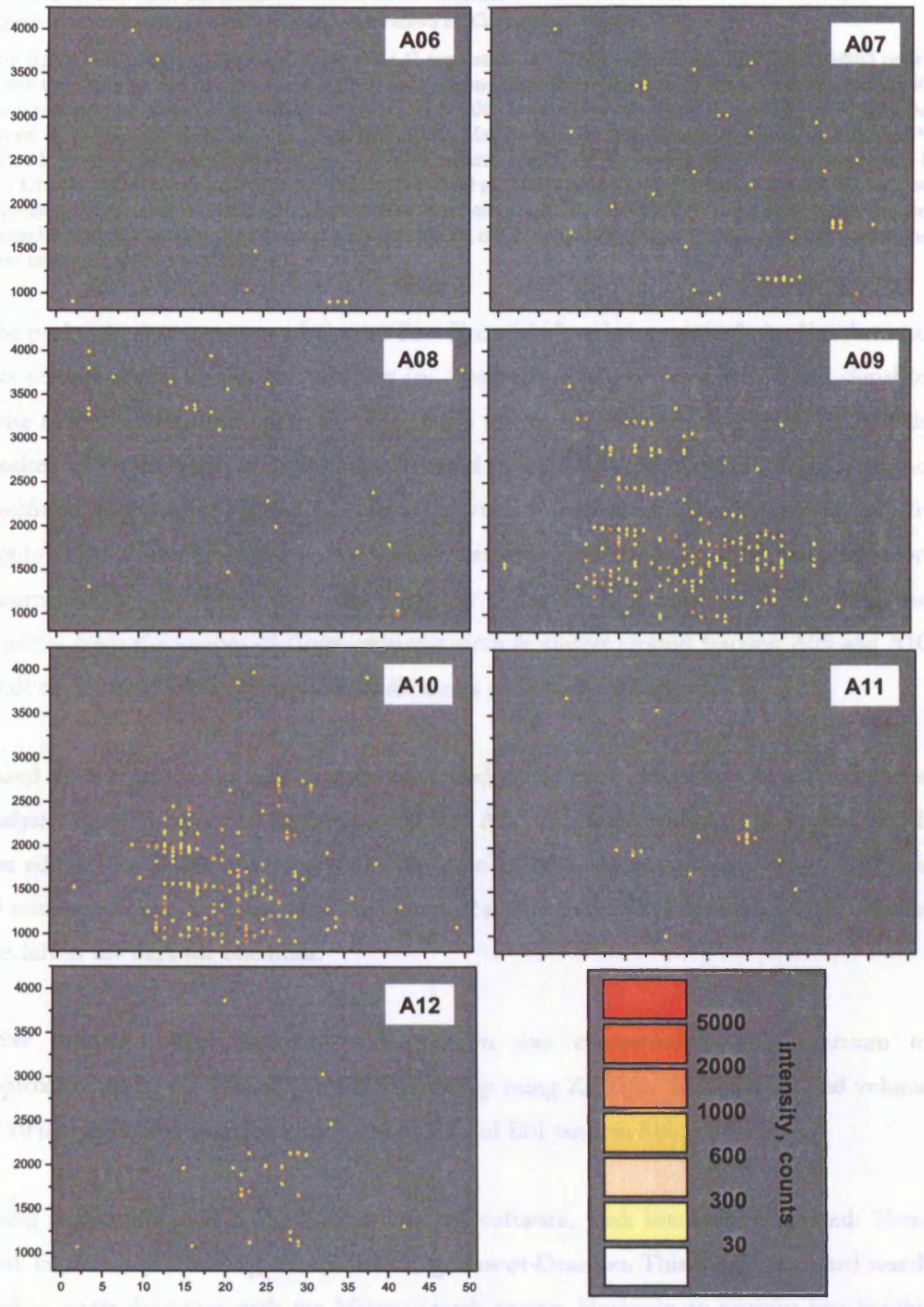

A06
A07
A08
A09
A10
A11
A12
4000
3500
3000
2500
2000
1500
1000
0
5
10
15
20
25
30
35
40
45
50
5000
2000
1000
600
300
30
intensity, counts

Figure 5.3.4 Mass maps of the 2D-LC separation of *Conus ventricosus*.

This figure shows the mass maps of the 2D-LC separation of *Conus ventricosus*. Each individual panel shows the mass to charge ratio measured of each second dimension fraction of *Conus ventricosus*. 1346 components were detected with S/N > 15 for m/z > 900. Intensity is displayed using the colour scheme shown in the bottom right hand panel of the figure. Matrix was spotted automatically using a Probot™ Micro Fraction Collector (LC Packings - Dionex). Matrix was made up from 1 part 70% methanol and 1 part CHCA. ACTH was added to the matrix (14 fmol·μl^{-1}) and used as an internal standard, during the automatic acquisition with the 4700 Proteomics Analyzer MALDI-TOF/TOF™. 1 µl analyte was added manually before allowing the mixture to air dry. Based on the results displayed in this figure, 96 fractions were taken for further analysis.

The peak lists that were recorded were first filtered before being plotted. An Excel macro was written which filtered the peak list for ions with m/z greater then 900 and signal to noise ratio over 15, identical to the mass maps shown for *Conus textile* (Figure 5.2.7). This resulted in the detection of 1346 ions, compared to the 1088 from *Conus textile*. This had no significant meaning as different amounts of venom were used. Also as can be seen in Figure 5.2.1 and Figure 5.3.1, the overall response (intensity) for the size exclusion chromatogram was higher for *Conus ventricosus* under identical conditions. However, the peptides from the venom of *Conus ventricosus* seem to cluster around fraction A09 and A10 while the overall elution from *Conus textile* seems to be more equally spread.

Based on the number of ions that were detected per fraction, 96 were selected for further analysis. Each fraction was buffered to 50 mM ABC. To 40 µl sample, 2 µl 20 mM TCEP was added. The solution was covered with argon and the reduction took place at 70 °C for 30 minutes. After a brief spin and cool down, 2 µl 200 mM IAM were added. The solution was left in the dark for one hour.

After reduction and alkylation, the solution was concentrated under vacuum to approximately 20 µl. The sample was cleaned up using ZipTips. This gave an end volume of 10 µl, which was analysed both by MALDI and ESI tandem MS.

Using scripts available in the Analyst analysis software, peak lists were generated. These peak lists were searched against NCBI using Mascot-Deamon. This is an automated search tool to query databases with the Mascot search engine. However no positive hits for this species was returned. To date 50 entries exists in the NCBI database for *Conus ventricosus*. Therefore all 96 LC-ESI-MS/MS runs were analysed manually. Figure 5.3.5 shows the ESI tandem mass spectrum from m/z 1156.922 $[M+2H]^{2+}$ from fraction A09-B11. From the

overlapping b-ion and y-ion series and internal fragment ions, 29 diagnostic ions were used to determine the sequence:

TVGDPERNYC PETCPTPCC m/z 2312.836 $[MH]^+$

The predominant ion in the spectrum was the unfragmented doubly charged parent ion. Some of the doubly charged b ions were also detected, these are labelled in purple in the figure. This sequence was searched against NCBI, using MS-Homology (PropeinProspector). The results from these searches revealed that this sequence was unique up to the point of 10 amino acid substitutions, which was over half of the peptide. This sequence at present could not be classified into a known superfamily (Terlau, 2004}(Cruz et al., 1985; Gray et al., 1988). The cysteine framework found for this sequence can be represented by -C-C-CC, putatively a new cysteine scaffold, or part of a degraded larger conotoxin. No further conotoxin like sequences were found in fraction A09-B11. The presence of arginine (R) in the sequence was a strong indication that the peptide was not subjected to the protease trypsin.

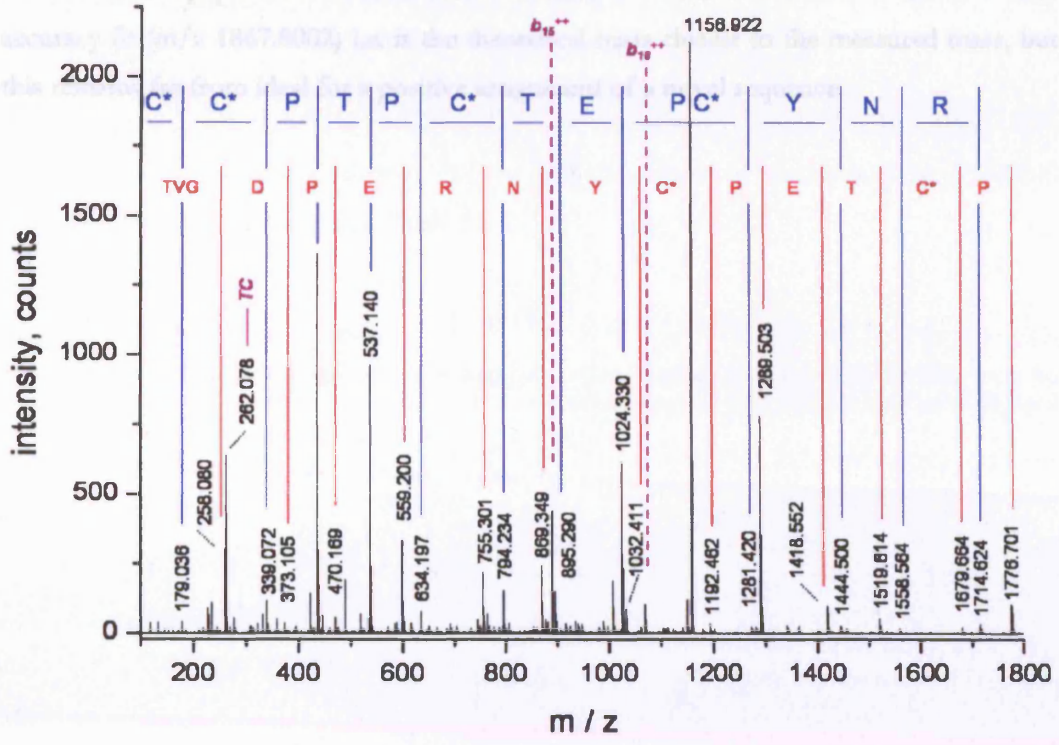


Figure 5.3.5 Tandem mass spectrum of m/z 1156.922 from fraction A09-B11 from Conus ventricosus.

This figure shows the annotated tandem mass spectrum of the doubly charged parent ion m/z 1156.922. This peptide (TVGDPERNYCPETCPTPCC) is found in fraction A09-B11. The near complete y-ion and b-ion series are annotated blue (top), and red (bottom) respectively. The most intense peak is unfragmented parent ion (m/z 1156.922). Also some doubly charged fragment ions are used to assign this sequence. Two of these are high lighted in dashed purple (b_{15} and b_{18}).

As is well documented, fragmentation after collision induced dissociation benefits greatly from the very basic guanidine group of arginine in the peptide backbone (Burlet et al., 1992; Csonka et al., 2004; Summerfield and Gaskell, 1997; Zhang, 2004). The tandem mass spectrum shown for fraction A09-B11 was a good example. Interpreting the fragmentation mass spectra of remaining fractions was a challenge. Often, the quality of fragmentation data, combined with the lack of an extensive *Conus* protein/peptide database gave only partial only sequence information. Figure 5.3.6 shows the tandem mass spectrum of m/z 934.404 $[M+2H]^{2+}$ from fraction A09-B06. The b-ion (red) and y-ion (blue) overlap at position b_3 / y_{14} .

SNVCA'xx'CYN GGQC'zz' m/z 1867.800 $[MH]^+$

Where 'xx' can be replaced by a number of dipeptide combinations, namely FH, QR or KR and 'zz' with dipeptide PP or GH. The combination 'KR' / 'PP' results in the best mass accuracy fit (m/z 1867.8002) i.e. is the theoretical mass closest to the measured mass, but this remains far from ideal for a positive assignment of a novel sequence.

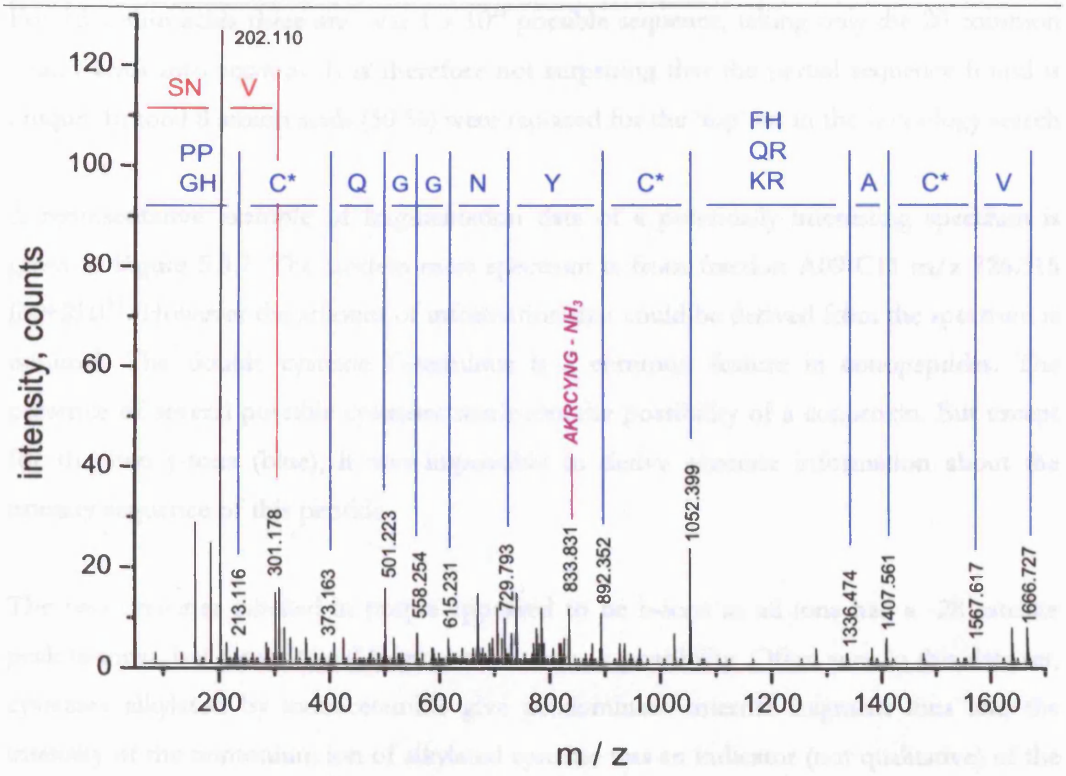


Figure 5.3.6 ESI tandem mass spectrum of m/z 934.404 from fraction A09-B06.

This figure shows the MS/MS spectrum of peptide SNVCAxxCYNGGQCxx (parent ion 934.399 $[M2H]^{2+}$). The y-ion series is displayed in blue. The b-ion series is displayed in red. The internal ion AKRCYNG – NH_3 is shown in purple. Multiple options are given when no positive assignment could be made.

The sequence SNVCAKRCYNGGQCPP was searched against NCBI using MS-Homology from ProteinProspector. For the query 8 amino acid substitutions were allowed. The 'top' hits are summarised in Table 5.3.1.

Table 5.3.1 Homology search results.

This table shows the top hit results of a homology search (MS-Homlogy from ProteinProspector) using the sequence SNVCA'xx'CYNGGQC'zz'. This peptide was searched against the NCBI database allowing for 8 substitutions.

Matching Sequence	Species	Protein Name
(K)IPLCAKSCVDDGQCPD(T)	Cyprinus Carpio	ovarian fibroin-like substance-2
(E et al.)TMVCNRHCENGGQCLT(E et al.)	Homo Sapiens	unnamed protein product
(E et al.)VNHCSNYCQNGGTCIP(S)	Mus Musculus	unnamed protein product

For 16 amino acids there are over 1 x 10^{24} possible sequence, taking only the 20 common amino acids into account. It is therefore not surprising that the partial sequence found is unique. In total 8 amino acids (50 %) were replaced for the 'top' hit in the homology search.

A representative example of fragmentation data of a potentially interesting spectrum is given in Figure 5.3.7. The tandem mass spectrum is from fraction A09-C11 m/z 726.315 $[M+2H]^{2+}$. However the amount of information that could be derived from the spectrum is minimal. The double cysteine C-terminus is a common feature in conopeptides. The presence of several possible cysteines reinforces the possibility of a conotoxin. But except for the two y-ions (blue), it was impossible to derive accurate information about the primary sequence of this peptide.

The two cysteines labelled in purple appeared to be b-ions as all ions had a -28 satellite peak (a-ions), but also internal fragment ions were a possibility. Often seen in this data set, cysteines alkylated by iodoacetamide give predominant internal fragment ions and the intensity of the immonium ion of alkylated cysteine was an indicator (not qualitative) of the number of cysteines present .

The ions labelled in black were not complementary ions of both the y-ions, b-ions or the expected internal fragment ions. The sum of the masses of the y-ion and its complementary b ion, with the same charge, equals that of the mass of the parent ion plus one proton. This is represented by the following equation:

$$y + b = p + H^+$$

y = y-ion $[MH]^+$ b = b-ion $[MH]^+$ p = parent ion $[MH]^+$

The survey scan revealed that a single ion was detected around the parent ion (m/z 726.3) (data not shown), minimising the possibility of a mixture of two peptide fragmentations spectra.

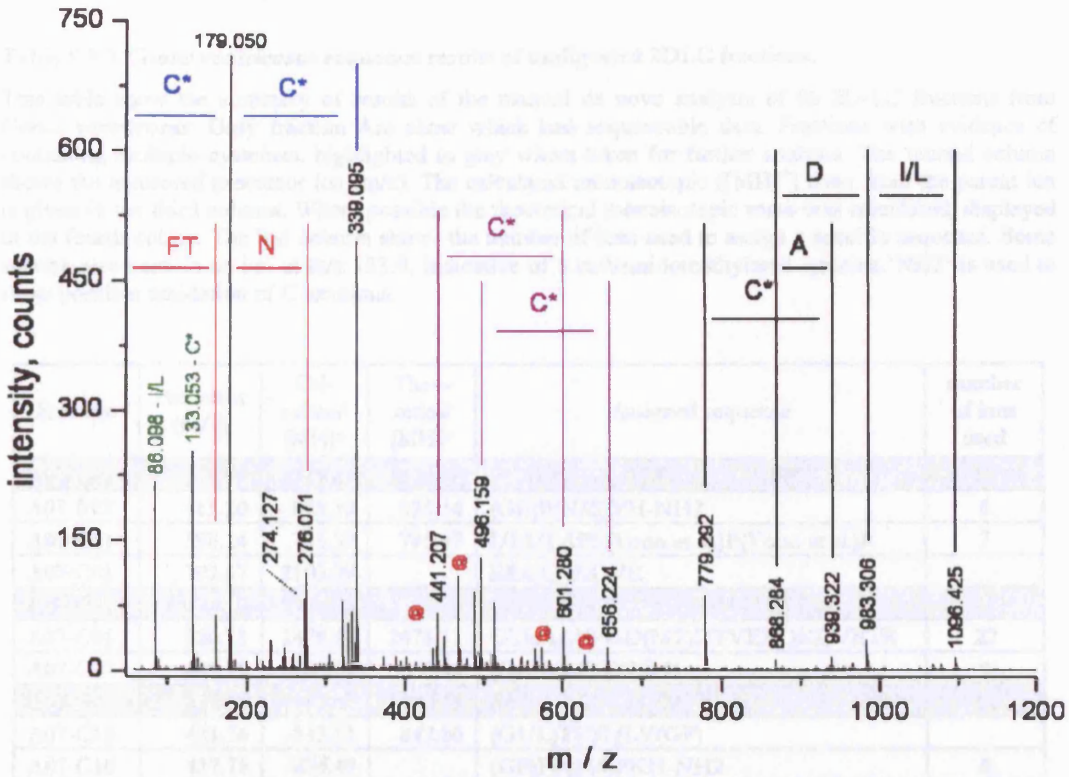


Figure 5.3.7 ESI tandem mass spectrum of m/z 726.315 from fraction A09-C11 from *Conus ventricosus*.

This figure is a representative example of a mass spectrum encountered during this analysis. The double cysteine in blue are y-ions. The double cysteine in purple appear to be b-ions as well as they all have -28 satellite peaks (indicated by red @) indicating loss of CO, indicative of b-ions. Labels in green are immonium ions.

As mentioned, the success rate of assigning unambiguous sequence was very low. The comprehensive analysis[4] of all 96 fractions is summarised in Table 5.3.2. for fractions which contained sequencable data, the precursor mass of the peptide in question is given. From the charge state the monoisotopic mass was calculated and where possible, a theoretical mass was generated using MS-Product (ProteinProspector). Complete and incomplete sequences are given, some of which contain gaps. In the last column the number of ions used to assign a specific sequence is given. Fractions which contained peptides with evidence of multiple cysteines, highlighted in grey, were selected for trypsin digestion. The results are shown in the next paragraph.

[4] Due to the overwhelming amount of data, some of de novo sequencing was performed by Dr. Craig Brinkworth (Mass Spectrometry Facility UCSF, California, USA)

Table 5.3.2 *Conus ventricosus* sequence results of undigested 2DLC fractions.

This table show the summary of results of the manual de novo analysis of 96 2D-LC fractions from *Conus ventricosus*. Only fraction Are show which had sequencable data. Fractions with evidence of containing multiple cysteines, highlighted in grey where taken for further analysis. The second column shows the measured precursor ion (m/z). The calculated monoisotopic ($[MH]^+$) mass from the parent ion is given in the third column. Where possible the theoretical monoisotopic mass was calculated, displayed in the fourth colmn. The last column shows the number of ions used to assign a specific sequence. Some spectra also contain an ion at m/z 133.0, indicative of a carbamidomethylated cysteine.'NH2' is used to show possible amidation of C terminus.

Fraction	Precursor (m/z)	Cal-culated $[MH]^+$	Theo-retical $[MH]^+$	Assigned sequence	number of ions used
A07-B09	776.95	2328.83	2328.97	(PC)VFNHVGEPCYyGI/LTPCC	15
A07-B12	413.20	825.39	825.34	(GE)P(NSS)VH-NH2	6
A07-C01	398.24	795.39	795.47	I/LI/LAPP(Yoon et al.)P(Yoon et al.)R	7
A07-C01	702.67	2105.99		EEQQ/KDVE	
A07-C01	974.10	2920.28		GLDD???KIVICCEGIVCTGLGR	26
A07-C06	826.73	2478.17	2478.13	GTEAAVPGD(MT)DTVEEQKDVETR	22
A07-C07	429.75	858.49	858.46	(GA)I/LEQVI/LE	9
A07-C07	567.28	1133.55	1133.45	(GA)PCSSNCN(PA)-NH2	11
A07-C10	421.76	842.51	842.50	(GI/L)TVSI/LV(GP)	
A07-C10	437.75	874.49		(GP)PP(146)PKH-NH2	8
A07-C10	481.27	961.53		(GP)PSP(146)PKH-NH2	8
A07-D02	648.71	1296.41	1296.47	GLCOWKPBC-NH2	7
A07-D02	700.87	1400.73		QPAPPI/LSVSAP	
A07-D02	911.42	1821.83		EVAAIVDNGS	17
A07-D02	765.72	2295.14	2295.09	EED(130.04)NELEEVKLESYPTIK	21
A07-D10	421.75	842.06	842.49	(GI/L)TVVTVPG	11
A07-D10	737.95	1474.90		TPI/LVGQPI/LSTP	13
A07-D11	421.76	842.06	842.49	(GI/L)TVVTVPG	13
A07-D11	929.80	2787.38	2787.39	(FD)GSDIITIDIPDIDPENEIIVSTR	35
A07-D12	719.70	2157.08	2157.02	AI/L(371.11)TTVAFGI/LTSEDWFK	18
A07-E01	1068.89	2136.77		(I/LC)CP????(I/LC)CV(PS)	
A09-B04	752.34	1503.67		DAGD(KG)AKPR	
A09-B06	549.78	1098.55		(GL)C?C?.....GDT	
A09-B06	650.31	1299.61	1299.65	I/LD(TA)QRADRI/L(EA)	15
A09-B06	934.39	1867.77	1867.80	SNVCAKRCYNGGQCPP	18
A09-B08	345.16	689.31	689.33	(DC)I/LPGK	5
A09-B08	413.22	825.43	825.37	VVTSSSMD	11
A09-B08	519.23	1037.45	1037.45	SSKNEFDpET-NH2	9
A09-B08	599.77	1198.53	1198.56	DSYVGDEAKSK	12
A09-B08	642.30	1283.59	1283.62	(PM)or(NN)or(DL)TAQPWRPEA	12
A09-B08	653.33	1305.65	1305.69	(DP)I/LI/LE(84)IHEK	
A09-B08	691.78	1382.55		(CA)IIICK	5
A09-B11	357.21	713.41	713.36	WVPSAPG	5
A09-B11	413.74	826.46	826.46	(EV)PIPSGK/Q	7
A09-B11	486.74	972.46		DPE(GT)NKVL or PEKQMKKVL	10
A09-B11	508.26	1015.51	1015.54	KPEDANI/LTK	8
A09-B11	515.28	1029.56	1029.56	APAVNTTEVK	23

A09-B11	606.29	1211.57	1211.58	PDEPPINETVT	12
A09-B11	680.82	1360.62	1360.62	AH(GT)ENEDGKFR	20
A09-B11	790.34	1579.66	1579.63	(MYN)PSAENEDGPQQ	16
A09-B11	1156.90	2312.83	2312.90	TVGDPERNYCPETCPTPCC	35
A09-C01	398.24	795.47	795.47	IIAPPER / IINM…	10
A09-C01	417.72	834.43	834.48	(VP)ADII(HA)-NH2	9
A09-C01	427.78	854.55		K?KAPAAAG??	
A09-C01	503.25	1005.49	1005.49	SSVDEAIER	12
A09-C01	547.25	1093.49	1093.49	PSAENEDGFK	18
A09-C01	611.28	1221.55		KGDENT	
A09-C01	627.84	1254.67		TTANIApEAI	
A09-C02	462.29	923.57		Q/KI/LI/LA	
A09-C02	489.20	977.39		(DP)D??PDL	
A09-C02	625.82	1250.63	1250.62	ISWPVDYTNR	9
A09-C06	509.71	1018.41		TPE(NG)A???L	
A09-C06	515.77	1030.53	1030.52	ADGDVEGVLR	12
A09-C06	518.80	1036.59	1036.58	VETGEPAPPR	13
A09-C06	545.24	1089.47		(PD)G(179.1)AGPFHL	11
A09-C06	582.30	1163.59	1163.60	(QI)IIAEDM(O)SK	10
A09-C06	594.31	1187.61	1187.60	EESAIQNAVAR	15
A09-C06	661.86	1322.71		VEVSAAI/LGTI/L?PR	12
A09-C06	442.59	1325.75		(DN)I/LQ/KGI/LTK??S-NH2	10
A09-C06	666.37	1331.73	1331.72	TTIEEEKINVR	16
A09-C06	689.33	1377.65	1377.68	RDFEKIDDSPR	21
A09-C06	691.82	1382.63	1382.62	EDSGI/LI/LEYDWR	14
A09-C06	722.37	1443.73		MQI/LENI/LQA??R	10
A09-C06	482.60	1445.78	1445.72	(SI/L)QTDAGEDE/LI/LQR	18
A09-C06	741.84	1482.67	1482.72	NFQPSAQVEQAQTQ	17
A09-C09	434.77	868.53	868.53	APVIDVVR	15
A09-C09	440.76	880.51	880.52	TPIPDIPK	15
A09-C09	466.25	931.49	931.51	ATAAAEITAL	12
A09-C09	491.82	982.63	982.60	APVIDRAIK	14
A09-C09	624.34	1247.67	1247.66	(QV)APVYSEVAGK	17
A09-C09	635.30	1269.59		DYGSDVSVVR	
A09-C09	636.33	1271.65	1271.65	(ID)EPTVVDEVR	11
A09-C09	661.34	1321.67	1321.64	SFDSVNPDALTR	20
A09-C09	668.81	1336.61		FPFSDISA(139)PES	9
A09-C09	710.90	1420.79		IPAVVVDTGT	14
A09-C09	735.35	1469.69		ESYTT	5
A09-C09	782.33	1563.65		NDFSGDFEEAA	10
A09-C11	513.80	1026.59		(I/LW) or (Q)VI/LSD	10
A09-C11	563.30	1125.59		EQTI/LRR	9
A09-C11	717.80	1434.59		C???CC	5
A09-C11	726.30	1451.59		I/L??C?C?CC	14
A09-C11	739.40	1477.59		NFES	6
A09-C11	760.80	1520.59		NF??I/LGNFES	16
A09-C11	774.40	1547.79		YAAEEFDGSF	15
A09-D01	818.40	1635.90		I/LNEI/LI/LEN	19
A09-D02	552.30	1103.59		PVI/LM	8
A09-D02	886.40	1771.79		(NH2-GN)QS(ETS) ????? DSI/LG	12
A09-D02	530.80	2120.08		SSSGSAV	11
A09-D03	439.22	877.43	877.43	EVTI/LDI/L(MG)	8

A09-D03	443.75	886.49	886.50	I/LENVI/LI/LW	7
A09-D03	452.26	903.51		I/LENVI/L?W	9
A09-D03	454.74	908.47		I/LEAVN	
A09-D03	472.26	943.51	943.53	(EG)IVNLLW or (AD)IVNLLW	8
A09-D03	478.31	955.61	955.56	GSPEIAIIR contains 133 peak	9
A09-D03	559.34	1117.67		(TV)GSI/L(130)I/LSI/LVV	
A09-D03	756.39	1511.77	1511.75	GYATGQFVVDEVAR	20
A09-D03	805.40	1609.80	1609.79	LGLDPIEEGVNYGPLH	17
A09-D03	886.39	1771.77	1771.79	(HA)AGI/LSDGSDFPSKDDGP-NH2	31
A09-D05	501.80	1002.59		??VI/LVRA-NH2	8
A09-D05	513.30	1025.59		NH2-(GI/L)GI/LSVTGI/LGG	16
A09-D05	518.80	1036.59		GI/LPA(NN)	11
A09-D05	640.80	1280.59		I/LEDQ	7
A09-D05	686.40	1371.79		SDI/LVGN	17
A09-D05	694.30	1387.59		FS??(F)PFSGD	14
A09-D06	646.40	1291.79		NH2-AGV??AEVWN(N)	13
A09-D07	614.30	1227.71		NH2-WPRDW	5
A09-D07	717.80	1434.49		ECC	4
A09-D07	496.50	1487.45		YGPC	4
A09-D07	774.40	1547.80		YAAEEFDGSF	13
A09-D09	722.90	1444.79		STYGI/LANI/LSA	15
A09-D09	771.90	1542.81		I/LI/LVK??YN	7
A09-D10	696.40	1391.79		??VVFTF-oh	6
A09-D10	819.90	1638.79		?I/LI/VAVE?	9
A09-D12	444.60	1332.79		??I/LQEEE?	9
A09-D12	492.00	1474.99		??(MN)I/L??SE	
A09-E01	453.20	905.39		GFAGDDAPR	9
A09-E01	771.60	2312.78		TPCC	4
A09-E01	1156.90	2312.78		TE??NR??T(PCC)	10
A09-E01	777.30	2329.88		TPCC	4
A09-E01	789.20	2365.58		TPCC	4
A09-E01	592.20	2365.78		TPCC	4
A09-E01	866.70	2597.80		PCC	3
A10-B09	347.15	693.29	693.36	AAAGEFK	12
A10-B09	357.18	713.35	713.35	ADVPSGAP	10
A10-B09	359.18	717.35	717.34	I/LASEGGSP	13
A10-B09	362.13	723.25	723.29	DCCIR	9
A10-B09	362.21	723.41	723.45	PPPVVR	13
A10-B09	364.66	728.31	728.33	YFGVQD	12
A10-B09	420.20	839.39	839.45	I/LVTSYEK	13
A10-B09	462.70	924.39		VSYAAEPST	
A10-B09	469.17	937.33	937.39	PFNDE(CO2)GSK	16
A10-B09	470.69	940.37	940.43	YGDEEISK	15
A10-B09	340.16	1018.46	1068.52	VGSDI/LSTSPR	20
A10-B09	525.72	1050.43	1050.49	DAPVDYTNR	15
A10-B10	771.60	2312.78		TPCC	4
A10-C01	699.00	2094.87		many Cysteines multiply charged	
A10-C02	447.70	894.39		RPI/LNV	11
A10-C04	787.90	2361.68		TPCC	4
A10-C12	534.80	1068.59		VQDFI/LD[185]	8
A10-C12	504.20	1510.58		I/LSTF	4
A10-C12	822.28	2464.85		CC	3

A10-D01	550.20	1099.45		GDSI/LGNET	
A10-D03	530.80	1060.59		ETI/LW	7
A10-D03	576.80	1152.59		I/LDAT	5
A11-C02	827.90	2481.68		(E et al.)Y?????SSCCY	11
A11-D01	822.60	2465.78		CSYESCCY ??? YPYA	17
A11-D01	1241.40	2481.79		CSY	4

5.3.2 Protease treatment of reduced and alkylated 2D-LC fractions of *Conus ventricosus*.

Peptide tandem mass spectra generated from low energy CID processes are often used for protein and peptide identification and characterisation (Medzihradszky et al., 2001; Wysocki et al., 2000). In a proteomic study, proteins are often digested with an enzyme to generate mass spectrometry amenable peptides. Trypsin is a highly specific serine protease which cleaves the amide bond on the C-terminus side of arginine (R) and lysine (K), except when followed by proline. The added advantage from the point of view of mass spectrometry is that tryptic fragments contain a basic amino acid at their C-terminus (R or K).

The fractions highlighted in grey in Table 5.3.2 were reduced, alkylated and digested with trypsin as described before. Briefly, a 250 µl fraction was concentrated to 100 µl, reduced with TCEP and alkylated with IAM. To approximately 70 µl, 20 ng of trypsin was added.

MALDI-TOF-MS spectra were recorded for (i) native (ii) reduced / alkylated (iii) and trypsin digested fractions, as shown in Figure 5.3.8. The native peptide (m/z 2116.6487 $[M+H]^+$) was alkylated six times by iodoacetamide. The mass increase by alkylation was 348.1562 Da, which indicated three disulfide bonds. When analysing the reduced / alkylated peptide, insource prompt fragmentation occurred. This is indicated by the red arrow in the figure and will be discussed in paragraph 5.3.3.

The peptide was hydrolysed by trypsin (bottom panel). The two proteolytic products were m/z 1014.3254 and m/z 1469.5177. The digestion reaction was incubated at 37°C for approximately 4 hours, however not all of the starting product was digested. Here also prompt fragmentation from the starting product was detected. All other labelled peaks were the result of deficient alkylation and the subsequent digestion of partially alkylated products.

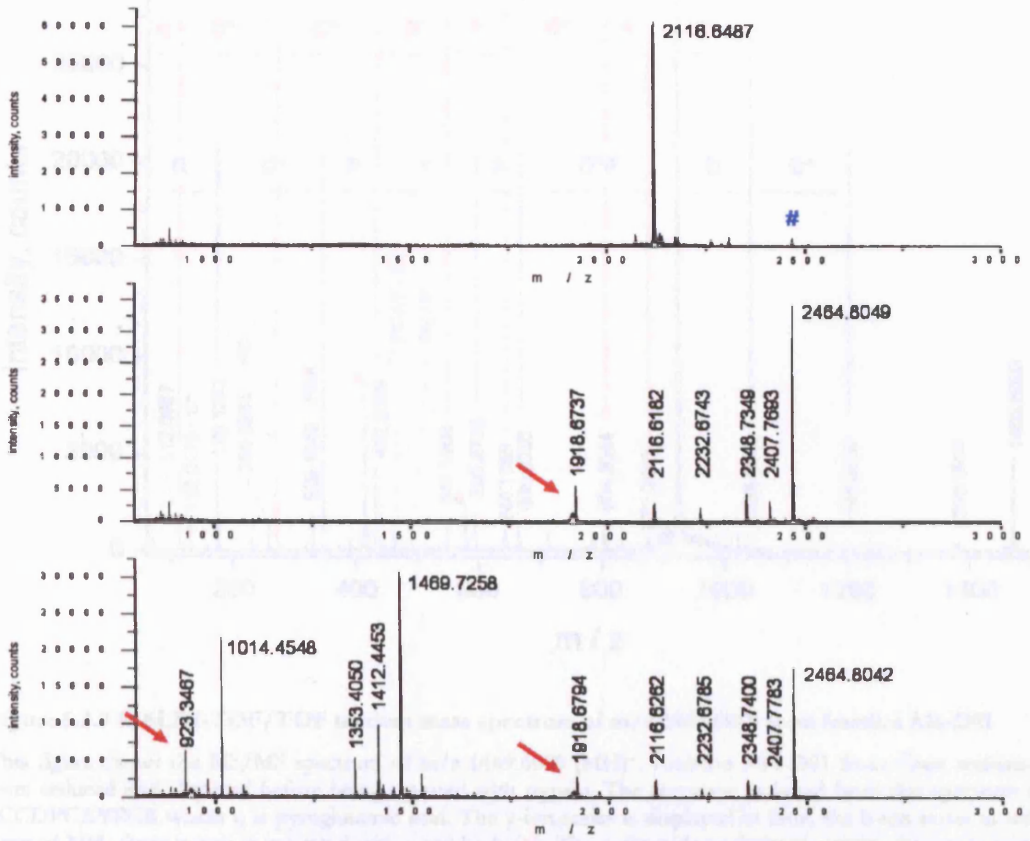


Figure 5.3.8 TOF-MS spectra of fraction A11-D01 from *Conus ventricosus*

This figure shows the MALDI-TOF-MS spectra of fraction A11-D01 from the 2D-LC separated *Conus ventricosus* venom. The top panel is the mass spectrum of the untreated fraction. The middle panel is the mass spectrum of the same fraction, but reduced (TCEP) and alkylated (IAM). The bottom panel shows the mass spectrum after trypsin digestion of the reduced and alkylated fraction. A theoretical digest shows that m/z 1014.3254 and m/z 1469.5177 result from a single cleavage by trypsin from 2464.8113. The ion indicated by the red arrow is the result of insource prompt fragmentation of the parent ion between aspartic acid and proline. See Figure 5.3.16 and Figure 5.3.17 for the zoom MS spectra and the M/MS spectra respectively.

Tandem mass spectra were recorded for the fully alkylated peptide, and for both the proteolytic products. This was done by LC-ESI-MS/MS and MALDI-TOF/TOF tandem MS. The next two figures (Figure 5.3.9 and Figure 5.3.10) show the MALDI tandem mass spectra from m/z 1469.6089 and m/z 1014.3666 (both singly charged) respectively. In the first spectrum an unambiguous sequence was assigned, namely

qCCDPCAYPCR m/z 1469.5070 $[MH]^+$

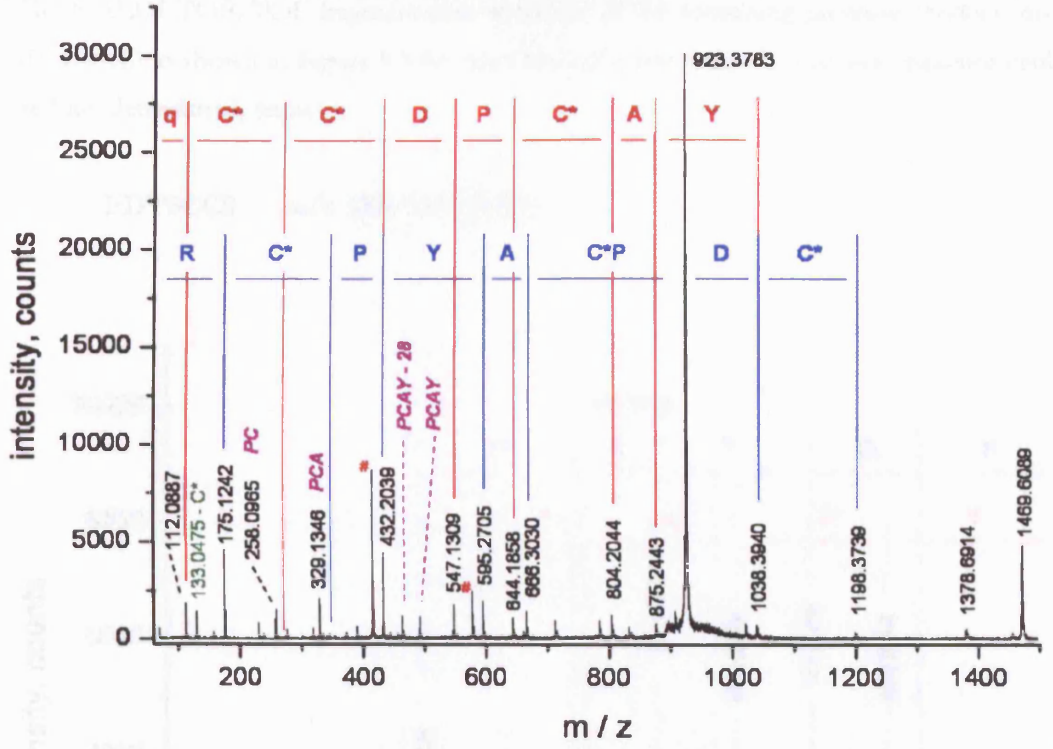


Figure 5.3.9 MALDI-TOF/TOF tandem mass spectrum of m/z 1469.6089 from fraction A11-D01

This figure shows the MS/MS spectrum of m/z 1469.6089 $[MH]^+$. Fraction A11-D01 from *Conus ventricosus* were reduced and alkylated before being digested with trypsin. The sequence assigned from this spectrum is qCCDPCAYPCR where q is pyroglutamic acid. The y-ion series is displayed in blue, the b-ion series in red. Loss of NH_3 from y-ions is indicated with a red hash (#). The carbamidomethylated cysteine immonium ion is shown in green. Internal fragment ions are displayed in purple.

Where 'q' is pyroglutamic acid (pyroGlu), the b_1 ion is detected at m/z 112.0887. PyroGlu is a natural amino acid and is, for instance, an important neurotransmitter in the vertebrate central nervous system. But PyroGlu acid can also be introduced by the experimental circumstances of the digestion, where the N-terminal Glu can cyclise to pyroGlu, resulting in the net loss of H_2O. However, this is not the case for this peptide, as the mass determined for the native peptide includes the pyroGlu modification. The major fragment ion resulted from the cleavage of the peptide backbone in between aspartic acid and proline. The charge was retained at the C-terminal side (y_7-ion). The fragmentation at this site is also the source of the prompt fragmentation seen in the MALDI-TOF-MS scans in Figure 5.3.8. Furthermore, some proline initiated internal fragments were detected, confirming parts of the sequence.

The MALDI-TOF/TOF fragmentation spectrum of the remaining protease product (m/z 1014.3668) in shown in Figure 5.3.10. Also from this spectrum the primary sequence could be fully determined, namely:

FDYSCCS m/z 1014.3337 $[MH]^+$

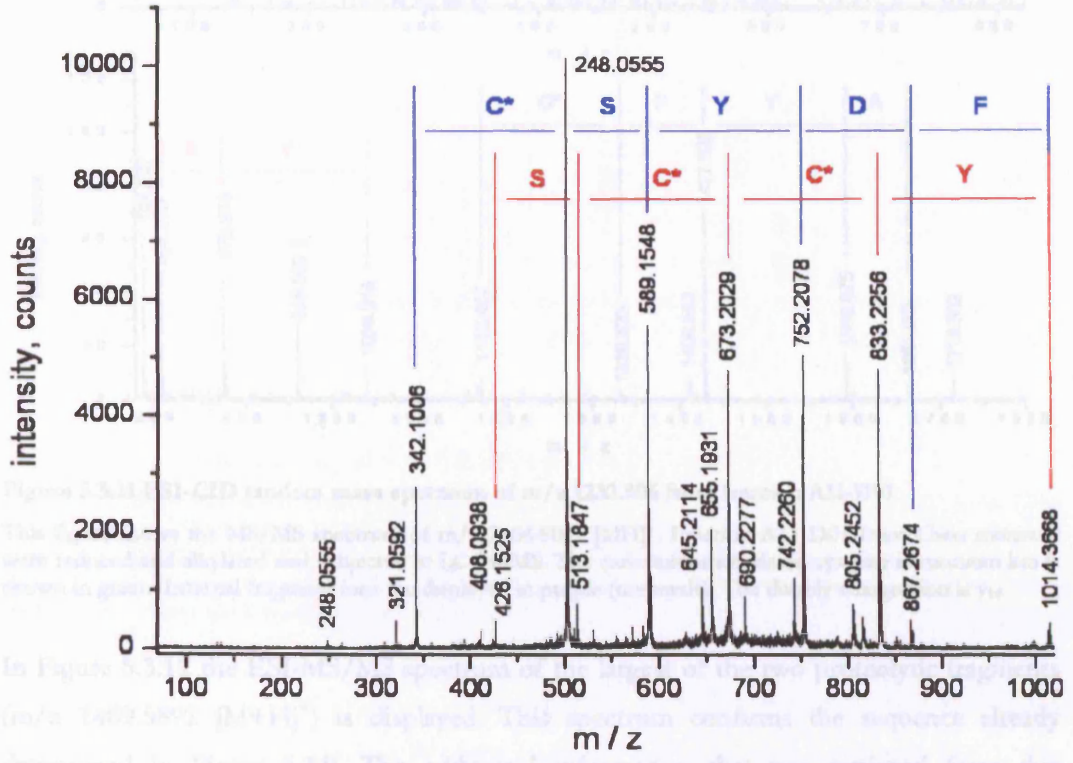


Figure 5.3.10 MALDI-TOF/TOF tandem mass spectrum of m/z 1014.3668

This figure displays the MS/MS spectrum of m/z 1014.3666 $[MH]^+$. Fraction A11-D01 from *Conus ventricosus* were reduced and alkylated before being digested with trypsin. The sequence assigned from this spectrum is FDYSCCS. The y-ion series is displayed in blue, the b-ion series in red.

These two sequences were verified by the analysis of the trypsin treated fraction by LC-ESI-MS/MS. The MS/MS spectrum of intact alkylated peptide is shown in Figure 5.3.11. The parent ion (m/z 1232.906) was doubly charged. With the knowledge of the previously determined sequences, many of the peaks can be assigned in this spectrum. As often, large peptides hardly ever lead to full assignment of the primary structure.

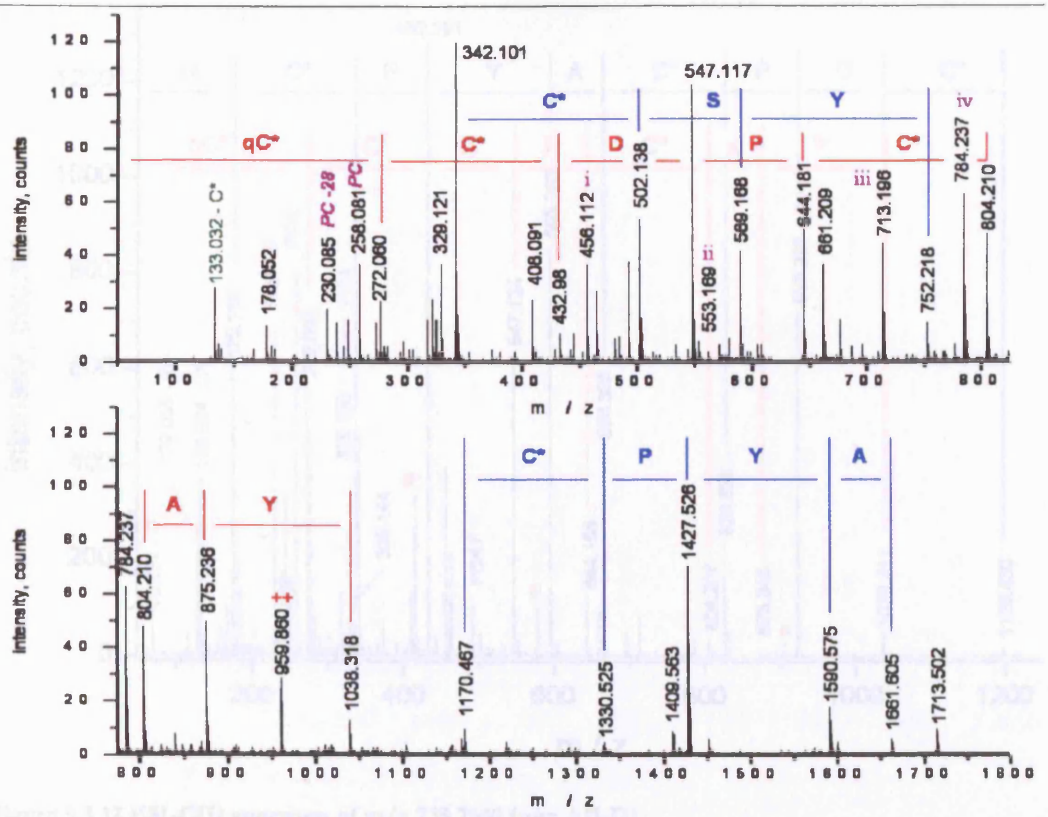


Figure 5.3.11 ESI-CID tandem mass spectrum of m/z 1232.906 from fraction A11-D01

This figure shows the MS/MS spectrum of m/z 2464.8042 $[MH]^+$. Fraction A11-D01 from *Conus ventricosus* were reduced and alkylated and subjected to LC-MSMS. The carbamidomethylated cysteine immonium ion is shown in green. Internal fragment ions are displayed in purple (numerals). The doubly charged ion is y_{14}.

In Figure 5.3.12 the ESI-MS/MS spectrum of the largest of the two proteolytic fragments (m/z 1469.5892 $[M+H]^+$) is displayed. This spectrum confirms the sequence already determined in Figure 5.3.9. The additional information that was retrieved from this spectrum was the confirmation of the amino acid sequence CDPC. In the MALDI fragment spectra the b_5-ion was not detected, unlike in the ESI fragment spectrum.

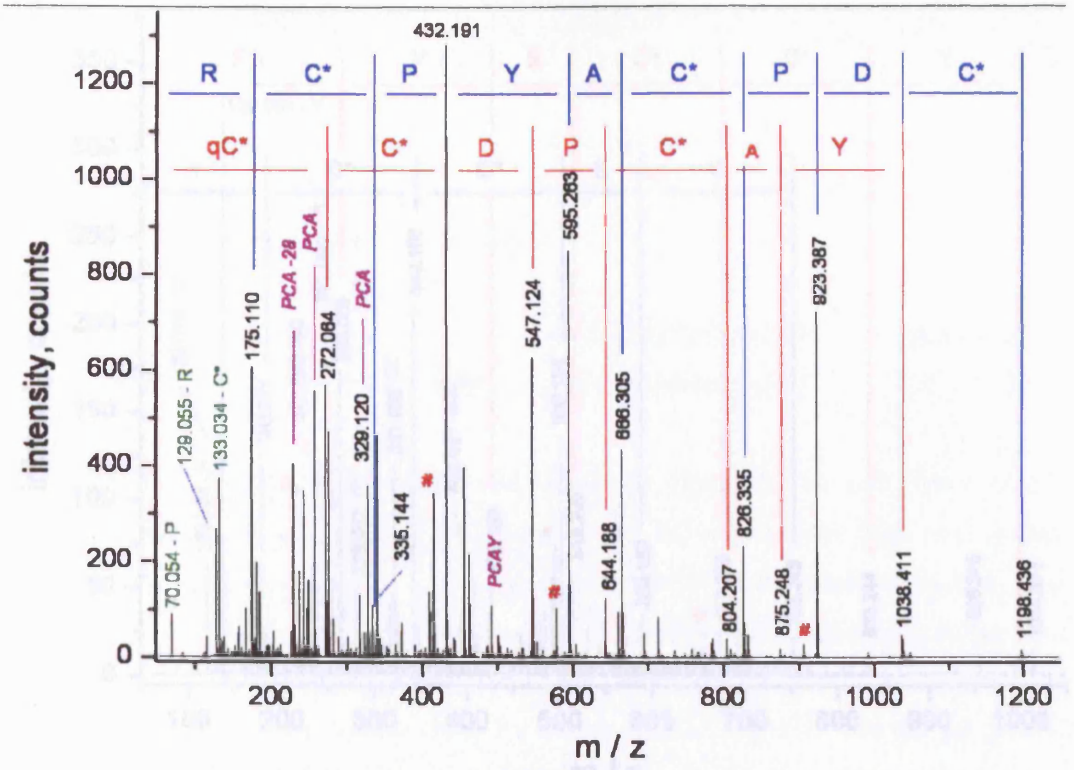


Figure 5.3.12 ESI-CID spectrum of m/z 735.7540 from A11-D1

This figure shows the tandem mass spectrum of m/z 1469.5892 $[MH]^+$ from fraction A11-D01 from *Conus ventricosus* venom. The sample was reduced and alkylated before being treated with trypsin. The assigned sequence is qCCDPCAYCPR where q is pyroglutamic acid. The y-ion series is shown in blue (top) and the b-ion series in red (bottom). Internal fragment ions are displayed in purple and immonium ions in green. Losses of NH_3 from y-ions are indicated with a red hash (#).

The opposite can be said about the next spectrum. In Figure 5.3.13, as in the previous spectrum, the sequence from the MALDI fragment spectrum was confirmed. But in this case, the assignment of FD would have been impossible without the MALDI tandem mass spectrum. Here the b_1-ion was not detected. The added advantage of ESI spectra is, that quite often, at least in the experience of the author, the immonium ions are an indication on the quantity of their accompanying amino acid. The predominant 136 and 133 immonium ions of tyrosine and carbamidomethylated cysteine respectively are good indicators when de novo sequencing an unknown peptide.

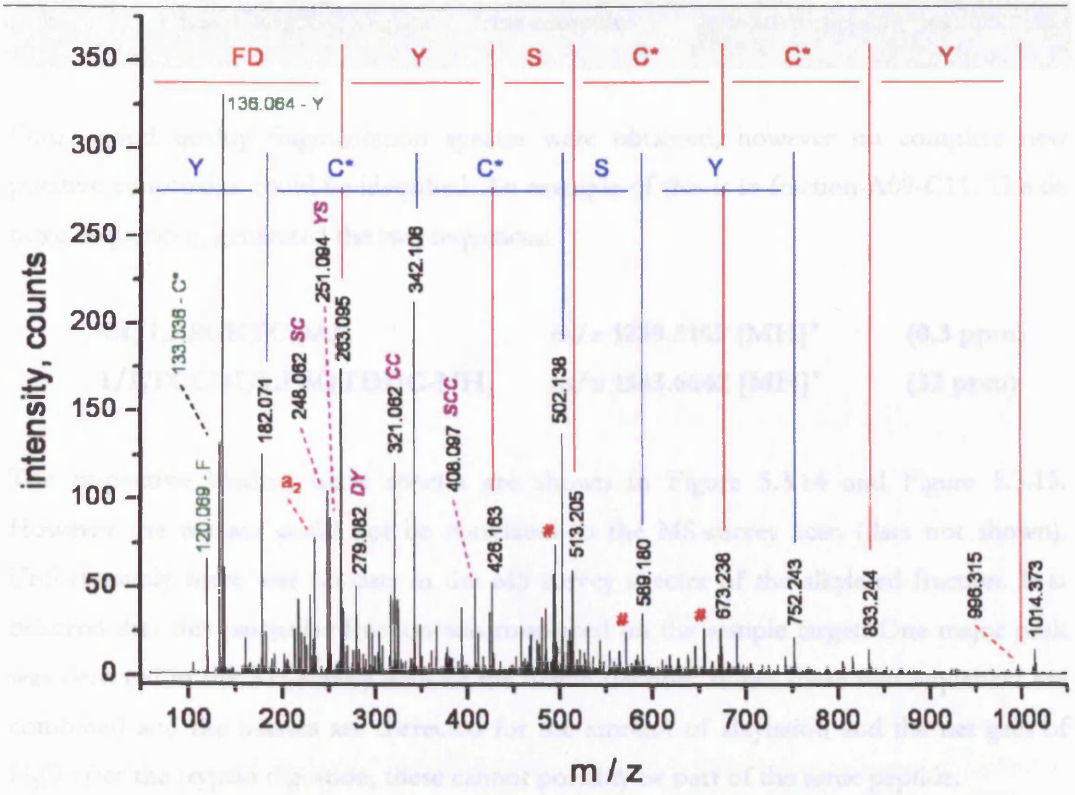


Figure 5.3.13 ESI-CID tandem mass spectrum of m/z 1014.373 of fraction A11-D1.

This figure shows the tandem mass spectrum of m/z 1014.373 $[MH]^+$ from fraction A11-D01 from *Conus ventricosus* venom. The sample was reduced and alkylated before being treated with trypsin. The assigned sequence is [FD]YSCCY. The y-ion series is shown in blue (top) and the b-ion series in red (bottom). Internal fragment ions are displayed in purple and immonium ions in green. Losses of H_2O from b-ions are indicated with a red hash (#).

The last five tandem mass spectra can be summarised by the next sequence, which prior to this writing has not been reported:

qCCDPCAYPCRFDYSCCY m/z 2464.8224 $[MH]^+$ (7 ppm)

This sequence is found in the NCBI database only when allowing for 8 substitution in an homology search (MS-Homology, ProteinProspector). A few examples are given in Table 5.3.3

Table 5.3.3 MS-Homology search results from *Conus ventricosus* fraction A11-D01

Matching Sequence	Species	Protein Name
(I)**VRCDHCAFPLRFDNACQY**(H)	Drosophila pseaudoobscure	GA22022-PA
(I)**VKCEHCSFPLRFDTSCQY**(H)	Drosphila melanogaster	RE38137p

(S)SCCQPCCRPSCCQSSCCK(E et al.)	Mus musculus	unnamed protein product

Other good quality fragmentation spectra were obtained, however no complete new putative conotoxins could be identified. An example of this is in fraction A09-C11. The de novo sequencing generated the two sequences:

CI/LASGETCDAR	**m/z 1239.5102 $[MH]^+$**	**(0.3 ppm)**
I/LTCCDI/LPSQTDDC-NH_2	**m/z 1583.6642 $[MH]^+$**	**(32 ppm)**

The respective tandem mass spectra are shown in Figure 5.3.14 and Figure 5.3.15. However the masses could not be correlated to the MS-survey scan (data not shown). Unfortunately there was no data in the MS survey spectra of the alkylated fraction. It is believed that this particular fraction was misplaced on the sample target. One major peak was detected in the MS-survey scan of the native fraction. When these two sequences are combined and the masses are corrected for the amount of alkylation and the net gain of H_2O after the trypsin digestion, these cannot possibly be part of the same peptide.

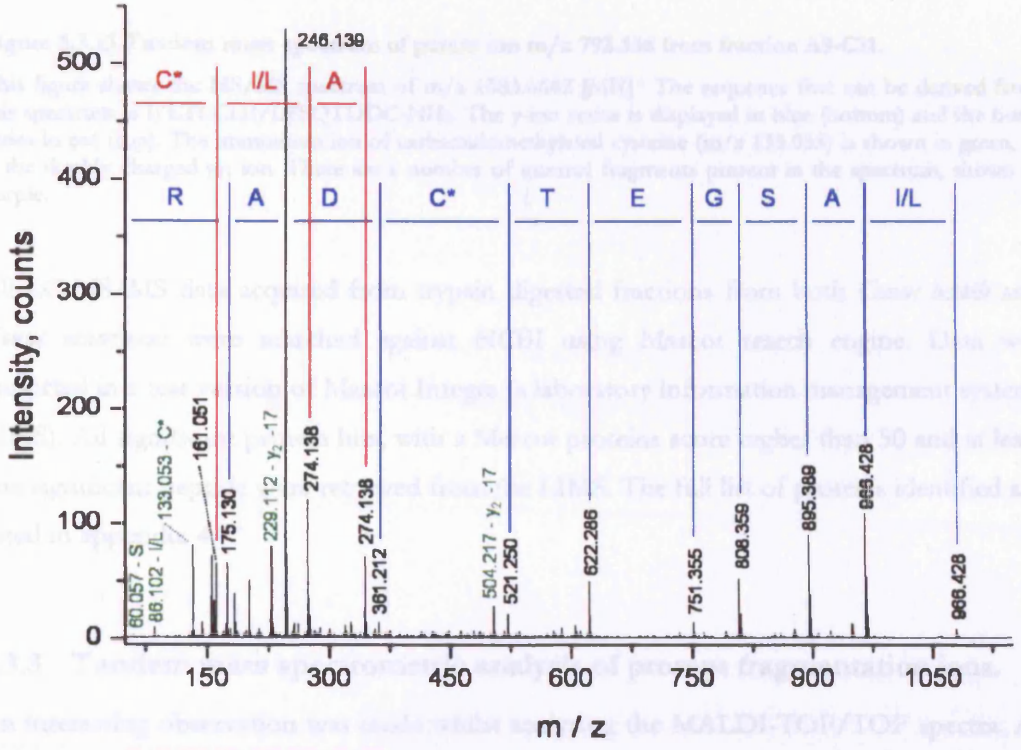


Figure 5.3.14 Tandem mass spectrum of m/z 620.259 of fraction A09-C11 from *Conus ventricosus*.

This figure shows the MS/MS spectrum of m/z 1239.5102 $[MH]^+$. CI/LASGETCDAR is assigned as the correct sequence. The y-ion series is displayed in blue, the b-ion series in red. Immonium ions and y-NH_3 ions are displayed in green.

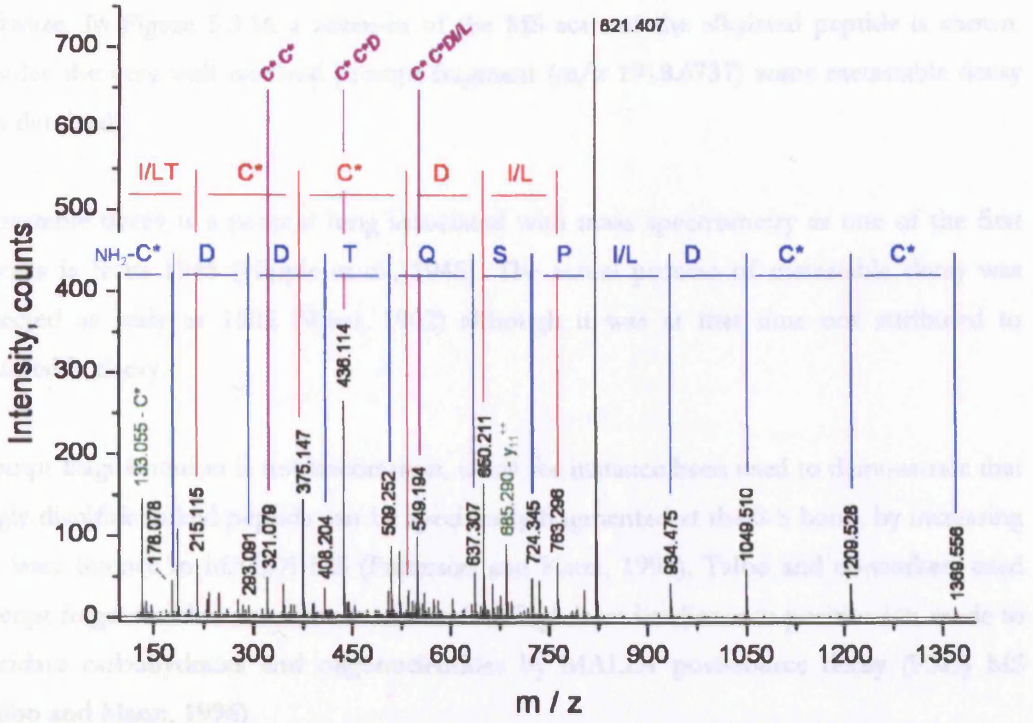


Figure 5.3.15 Tandem mass spectrum of parent ion m/z 792.336 from fraction A9-C11.

This figure shows the MS/MS spectrum of m/z 1583.6642 $[MH]^+$. The sequence that can be derived from this spectrum is I/LTCCDI/LPSQTDDC-NH_2. The y-ion series is displayed in blue (bottom) and the b-ion series in red (top). The immonium ion of carbamidomethylated cysteine (m/z 133.055) is shown in green, as is the doubly charged y_{11} ion. There are a number of internal fragments present in the spectrum, shown in purple.

All LC-MS/MS data acquired from trypsin digested fractions from both *Conus textile* and *Conus ventricosus* were searched against NCBI using Mascot search engine. Data was imported in a test version of Mascot Integra (a laboratory information management system, LIMS). All significant protein hits, with a Mascot proteins score higher than 50 and at least one significant peptide were retrieved from the LIMS. The full list of proteins identified are listed in appendix 4.

5.3.3 Tandem mass spectrometric analysis of prompt fragmentation ions.

An interesting observation was made whilst analysing the MALDI-TOF/TOF spectra. As mentioned before, the MALDI desorption/ionization process of the alkylated peptide

from fraction A11-D01 created prompt fragmentation. This form of in-source fragmentation occurs in the ionization chamber simultaneously with or immediately after ionization. The prompt fragment was selected for MS/MS by the automatic acquisition software. In Figure 5.3.16 a zoom-in of the MS scan of the alkylated peptide is shown. Besides the very well resolved prompt fragment (m/z 1918.6737) some metastable decay was detected.

Metastable decay is a process long associated with mass spectrometry as one of the first reports is from 1945 (Hipple et al., 1945). The actual process of metastable decay was detected as early as 1902 (Wien, 1902) although it was at that time not attributed to metastable decay.

Prompt fragmentation is not uncommon, it has for instance been used to demonstrate that singly disulfide linked peptide can be specifically fragmented at the S-S bond, by increasing the laser fluence in MALDI-MS (Patterson and Katta, 1994). Talbo and co-workers used prompt fragmentation to remove sialic acids at high laser irradiance in positive-ion mode to elucidate carbohydrates and oligonucleotides by MALDI post-source decay (PSD) MS (Talbo and Mann, 1996).

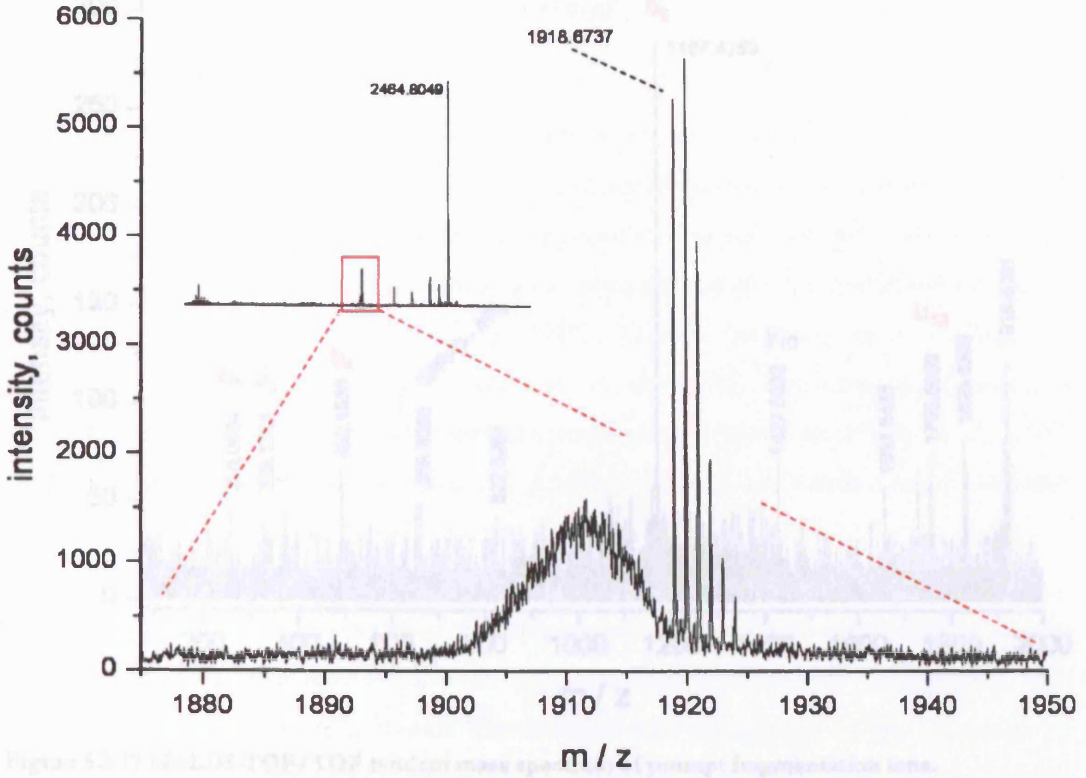


Figure 5.3.16 Prompt fragmentation of peptides between Aspartic acid and Proline.

This figure shows a zoom-in of the MS survey scan from the alkylated peptide from fraction A11-D01. Two of the in source fragmentation processes are visible, besides well resolved prompt fragmentation, metastable decay was detected. The prompt fragmentation occurred between the aspartic acid (D) and proline (E et al.).

The tandem mass spectrum is shown in Figure 5.3.17. Although the intensity of most of the fragments was low, 5 b-ions and 1 y-ion could be assigned as well as an internal fragment ion. Even though it is unlikely that an entire primary sequence can be determined, the information was used to confirm sequences determined from fragmentation of ordinary parent ions. The fragmentation of this prompt fragment also follows the pattern of preferable cleavage of aspartic acid on the C-terminal side of the amino acid. In Figure 5.3.18 the fragments that were detected are represented in a schematic.

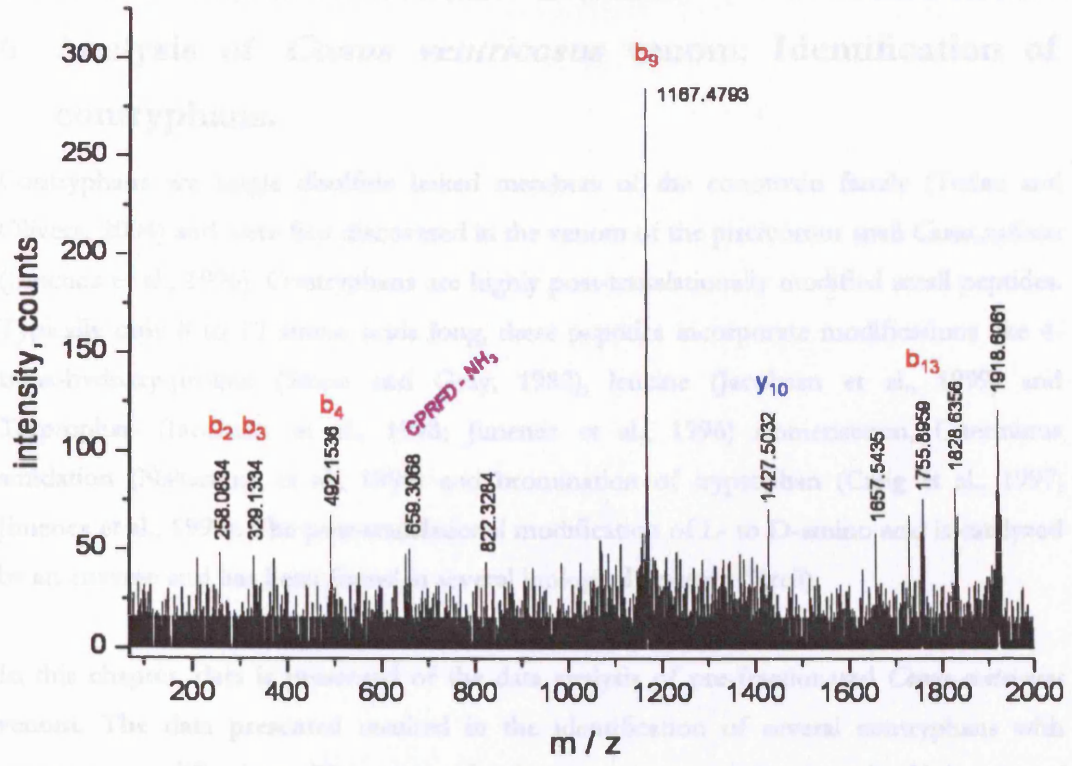


Figure 5.3.17 MALDI-TOF/TOF tandem mass spectrum of prompt fragmentation ions.

The tandem mass spectrum of the prompt fragment PCAYCPRFDYSCCY is shown in this figure. The prompt fragment was created after the laser desorption/ionization process of the alkylated peptide qCCDPCAYPCRFDYSCCY.

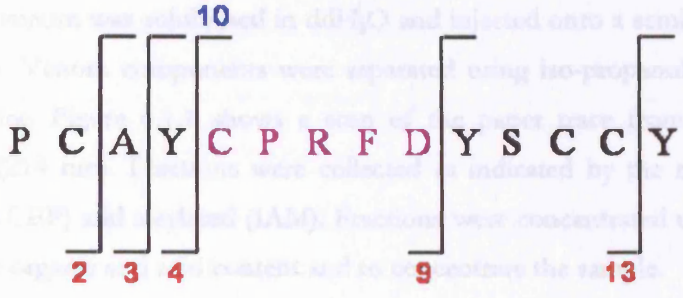


Figure 5.3.18 Fragmentation scheme of m/z 1918.6061.

The detected ions in Figure 5.3.17 are represented in this figure by a fragmentation scheme. The b-ions are labelled in red, the y-ion in blue. The part of the peptide backbone that was detected as internal fragment is displayed in purple.

6 Analysis of *Conus ventricosus* venom: Identification of contryphans.

Contryphans are single disulfide linked members of the conotoxin family (Terlau and Olivera, 2004) and were first discovered in the venom of the piscivorous snail *Conus radiatus* (Jimenez et al., 1996). Contryphans are highly post-translationally modified small peptides. Typically only 8 to 12 amino acids long, these peptides incorporate modifications like 4-trans-hydroxy-proline (Stone and Gray, 1982), leucine (Jacobsen et al., 1999) and Tryptophan (Jacobsen et al., 1998; Jimenez et al., 1996) isomerisation, C-terminus amidation (Nakamura et al., 1996) and bromination of tryptophan (Craig et al., 1997; Jimenez et al., 1997). The post-translational modification of L- to D-amino acid is catalyzed by an enzyme and has been found in several biological systems (Kreil).

In this chapter, data is presented of the data analysis of pre-fractionated *Conus ventricosus* venom. The data presented resulted in the identification of several contryphans with numerous modifications. The venom fractionation was carried out at the University of California San Francisco, Mass Spectrometry facility (Prof. Dr. Alma L. Burlingame) by Dr. Kirk C. Hansen.

6.1 Venom preparation and separation

Conus ventricosus venom was solubilised in ddH_2O and injected onto a semi-preparative C18 column (Vydac). Venom components were separated using iso-propanol (2-propanol) as the organic phase. Figure 6.1.1 shows a scan of the paper trace from the absorbance chromatogram (214 nm). Fractions were collected as indicated by the red numbers and were reduced (TCEP) and alkylated (IAM). Fractions were concentrated under vacuum, in order to remove organic and acid content and to concentrate the sample.

The resulting samples were re-injected onto a capillary C18 column as in previous chapters. In order to create a semi orthogonal separation, acetonitrile (instead of iso-propanol in the first dimension) was used as the organic phase in this second dimension.

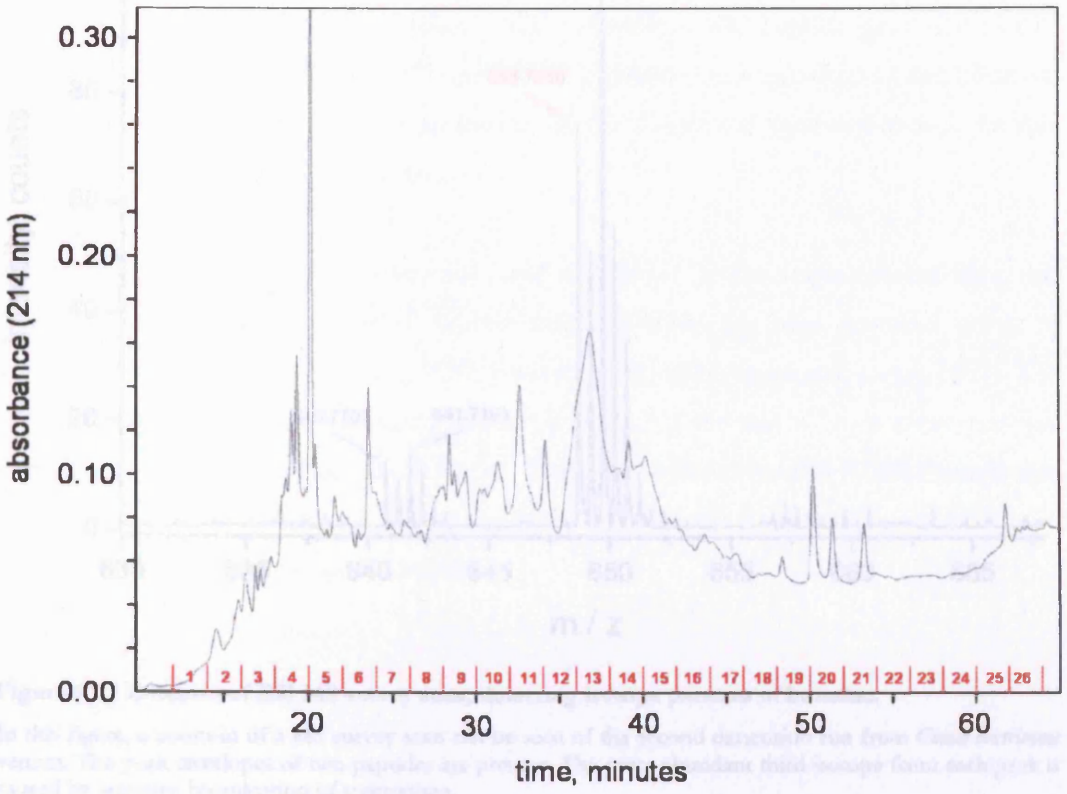


Figure 6.1.1 Reversed phase chromatogram of *Conus ventricosus* venom.

This figure shows the UV absorbance (214 nm) of the separation of *Conus ventricosus* venom. Venom was solubilised in ddH_2O and separated on a semi-preparative C18 reversed phase column, using iso-propanol as the organic phase. Fractions were taken as indicated in red. A number of these fractions were further analysed by LC-ESI-MS/MS with ACN as the organic phase.

6.2 2D LC analysis of Conus ventricosus venom.

Figure 6.2.1 shows the ESI MS-survey scan of two co-eluting peptides. As the isotope distribution is clearly remarkable, the third and following isotopes are clearly out of proportion when compared to a standard "averagine" peptide. Averagine is used to calculate isotope ratios, using an average amino acid (Johnson and Muddiman, 2004). Such isotope distributions were seen in case of partially reduced peptides, where for each disulfide formed or broken a 2.0156 Dalton differential peptide is formed. However, when analysing the tandem mass spectra, a complete different picture unfolded.

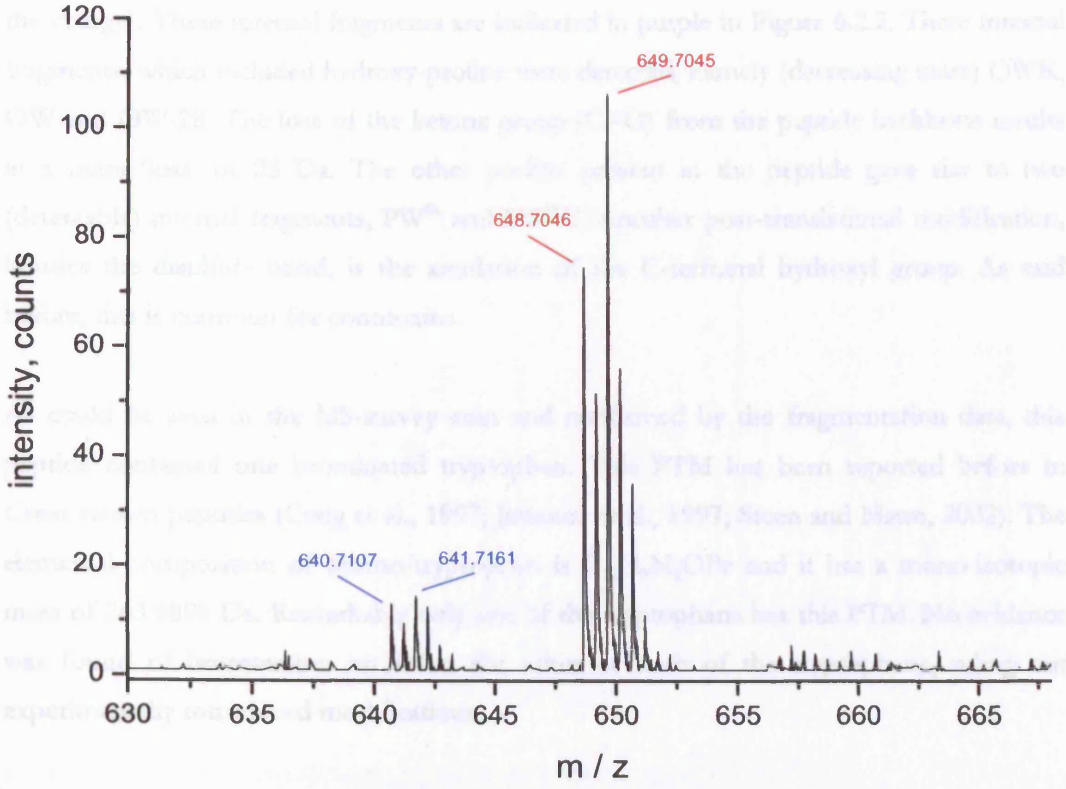


Figure 6.2.1 Zoom-in of ESI MS survey scan, detecting isotope patterns of bromine.

In this figure, a zoom-in of a MS survey scan can be seen of the second dimension run from *Conus ventricosus* venom. The peak envelopes of two peptides are present. The more abundant third isotope from each peak is caused by selective bromination of tryptophan.

The tandem mass spectrum from the more intense ion (m/z 648.7046) from Figure 6.2.1 is shown is Figure 6.2.2. The rather exotic sequence determined by tandem MS from this fragment spectrum is:

G(I/L)COWKPW^{Br}C-NH$_2$ m/z 1296.4014 [MH]$^+$ (49 ppm)

Where "O" is hydroxy-proline and "W^{Br}" is bromo-tryptophan. Cysteines (E et al.) are modified by acetamide. The C-terminus is amidated (NH_2). The measured parent mass is within 50 ppm of the theoretical mass.

Hydroxy-proline ($C_5H_7NO_2$) is a rare post-translational modification. Because its mass is near to that of (iso)leucine ($C_6H_{11}NO$) with a difference of 0.03636 Da, it is easily mistaken during de novo sequencing for the latter. However, hydroxy-proline behaves in a similar

way as proline regarding its fragmentation characteristics. Both give rise to prominent internal fragments (the backbone is cleaved in two places and the internal fragment retain the charge). These internal fragments are indicated in purple in Figure 6.2.2. Three internal fragments, which included hydroxy-proline were detected, namely (decreasing mass) OWK, OW and OW-28. The loss of the ketone group (C=O) from the peptide backbone results in a mass 'loss' of 28 Da. The other proline present in the peptide gave rise to two (detectable) internal fragments, PW^{Br} and $PW^{Br}C$. Another post-translational modification, besides the disulfide bond, is the amidation of the C-terminal hydroxyl group. As said before, this is common for conotoxins.

As could be seen in the MS-survey scan and confirmed by the fragmentation data, this peptide contained one brominated tryptophan. This PTM has been reported before in *Conus* venom peptides (Craig et al., 1997; Jimenez et al., 1997; Steen and Mann, 2002). The elemental composition of bromo-tryptophan is $C_{11}H_9N_2OBr$ and it has a mono-isotopic mass of 263.9898 Da. Remarkably only one of the tryptophans has this PTM. No evidence was found of bromination on either the other or both of the tryptophans, ruling out experimentally introduced modifications.

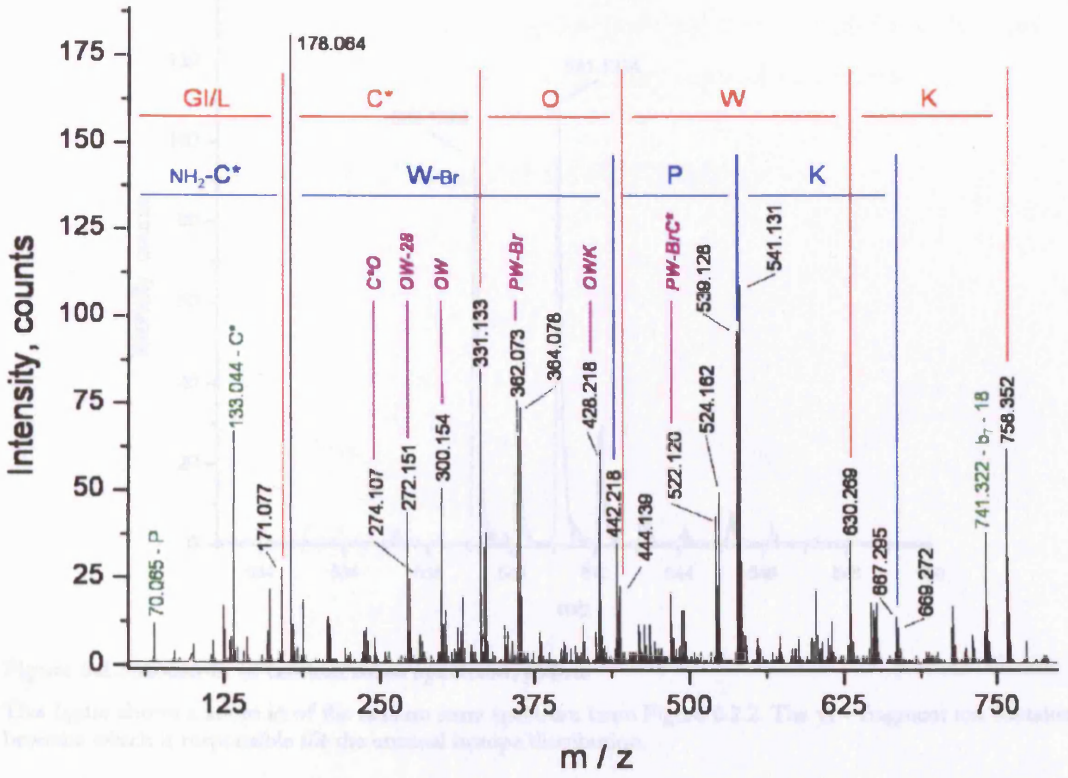


Figure 6.2.2 Tandem mass spectrum of bromo-contryphan m/z 648.7046

This figure shows the MS/MS mass spectrum of m/z 1296.4014 $[MH]^+$. The y-ions series is shown in blue, while the b-ion series is shown in red. Internal fragment ions are displayed in purple. For y ions 2,3 and 4 also the third isotope mass is labelled. The distorted isotope pattern is caused by the incorporation of bromine on Tryptophan.

The unusual isotope distribution was also detected in the tandem mass spectrum of the fragments containing the modified tryptophan (y_2, y_3 and y_4). A zoom in of the raw spectrum for the y_3 fragment ion (m/z 539.126) is shown in Figure 6.2.3. the intense third isotope (m/z 541.1314) is clearly visible.

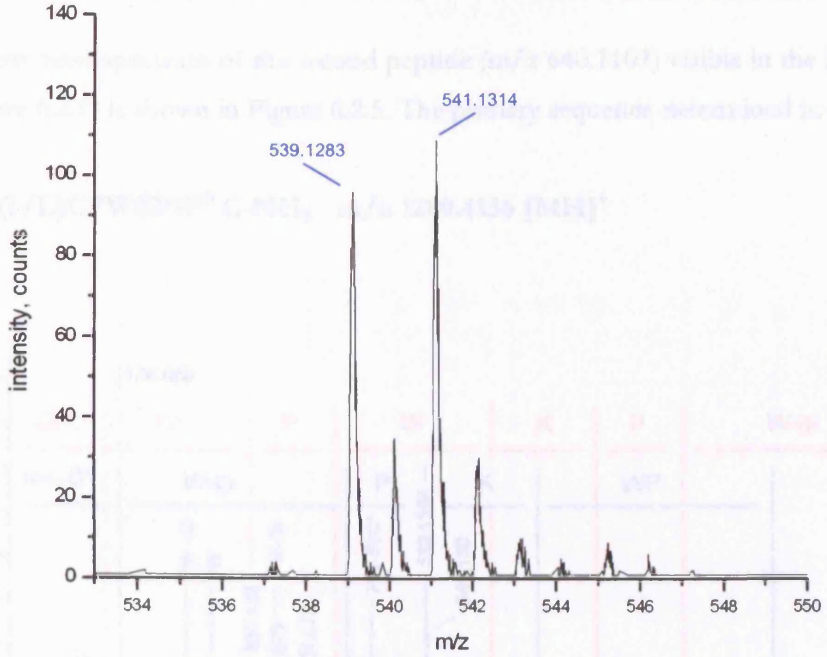
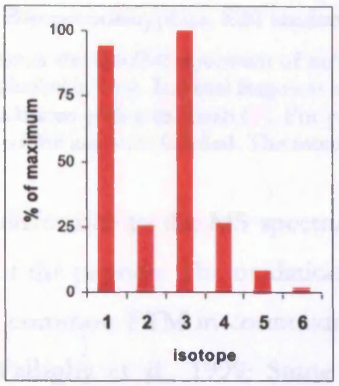


Figure 6.2.3 Zoom-in of tandem mass spectrum, y_3-ion.

This figure shows a zoom in of the tandem mass spectrum from Figure 6.2.2. The y_3 – fragment ion contains bromine which is responsible for the unusual isotope distribution.

The theoretical isotope distribution of the y_3 fragment ion was calculated with MS-Isotope, the software package of ProteinProstector (Clauser et al., 1999).The calculated masses of the y_3 ion, with the elemental composition $C_{21}H_{28}N_6O_4BrS$ and the graphical representation are displayed in Figure 6.2.4. The experimental data matches the calculated values. This was also the case for the other y-ions that incorporated bromine (data not shown).



y_3 : P – W(Br) – C* – NH_2

isotope number	m/z	percent total	percent maximum
1	539.1076	36.8	94.3
2	540.1104	10.0	25.5
3	541.1058	39.0	100.0
4	542.1084	10.3	26.4
5	543.1062	3.2	8.3
6	544.1070	0.6	1.6

Figure 6.2.4 Theoretical isotope distribution of bromo-tryptophan containing y_3-ion.

The panel on the left hand side shows the graphical representation of the theoretical isotope distribution of the y_3 fragment ion extracted from the table on the right hand side of the figure. The theoretical isotope distribution was calculated with the MS-Isotope feature of ProteinProspector using the elemental composition $C_{21}H_{28}N_6O_4BrS$.

The tandem mass spectrum of the second peptide (m/z 640.7107) visible in the MS-survey scan (Figure 6.2.1) is shown in Figure 6.2.5. The primary sequence determined is:

G(I/L)CPWKPW^{Br} C-NH$_2$ m/z 1280.4136 [MH]$^+$

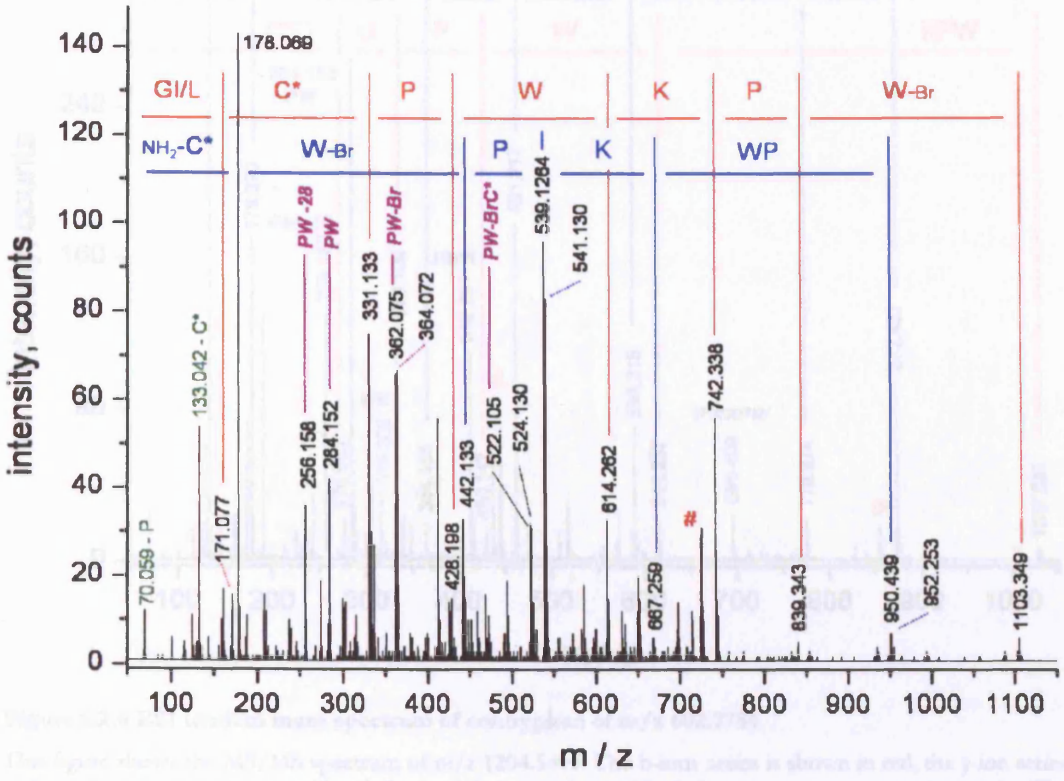


Figure 6.2.5 Bromo-contryphan, ESI tandem mass spectrum of m/z 640.7107

This figure shows the MS/MS spectrum of m/z 1280.4136 [MH]$^+$. The b-ion series is labelled in red. The y-ion series is labelled in blue. Internal fragment ions are shown in purple, immonium ions in green and the loss of water from b-ions with a red hash (#). For y-ions containing the brominated tryptophan, both the first and third isotope of the ion were labelled. The same for internal fragment ions.

The mass difference in the MS spectrum, 7.99 Da (15.9878 for singly charged), indicated oxidation of the peptide. The oxidation site is on Proline in position 4. Hydroxy-proline is a relatively common PTM in conotoxins (Cruz et al., 1985; Fainzilber et al., 1995a; Hill et al., 1996; Pallaghy et al., 1999; Stone and Gray, 1982). In a similar fashion to bromo-tryptophan, evidence was found that supported the oxidation of this particular hydroxy-proline only. The internal fragments (purple labels) detected in this fragment data originated from both prolines present in the peptide.

Compared with the previous tandem mass spectra, more ions were detected in the high mass range (truncated for clarity in Figure 6.2.2). The b-ion series overlaps the y-ion series for this particular peptide.

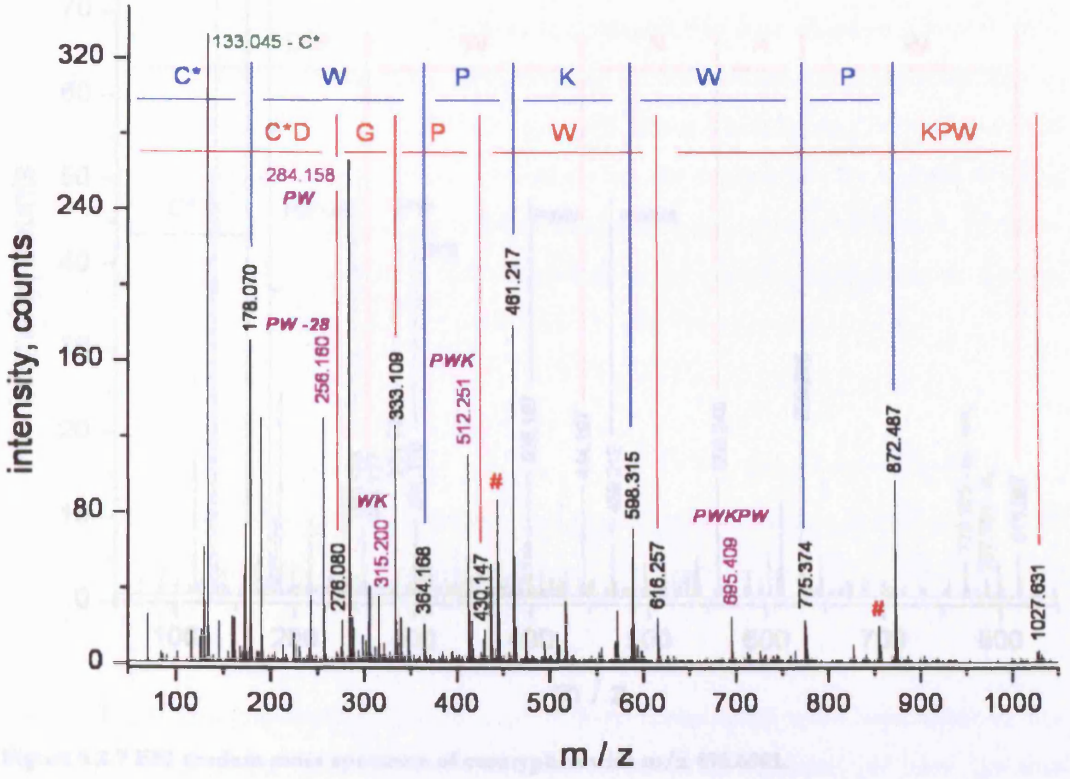


Figure 6.2.6 ESI tandem mass spectrum of contryphan of m/z 602.7784

This figure shows the MS/MS spectrum of m/z 1204.5496. The b-ions series is shown in red, the y-ion series in blue. The C-terminus is amidated. Internal fragment ions with their sequence are shown in purple. The losses of ammonium from y-ions are indicated with a red hash (#). The immonium ion is labelled in green.

Another two contryphans were found in the remaining second dimension fractions. The contryphan with sequence

CDGPWKPWC-NH_2 m/z 1204.5496 $[MH]^+$ (38 ppm)

Is shown in Figure 6.2.6. And the contryphan with sequence:

CPWNAWC-NH_2 m/z 992.1922 $[MH]^+$ (31 ppm)

Is shown in Figure 6.2.7. None of the contryphans shown in this chapter have been, to date, reported in the literature.

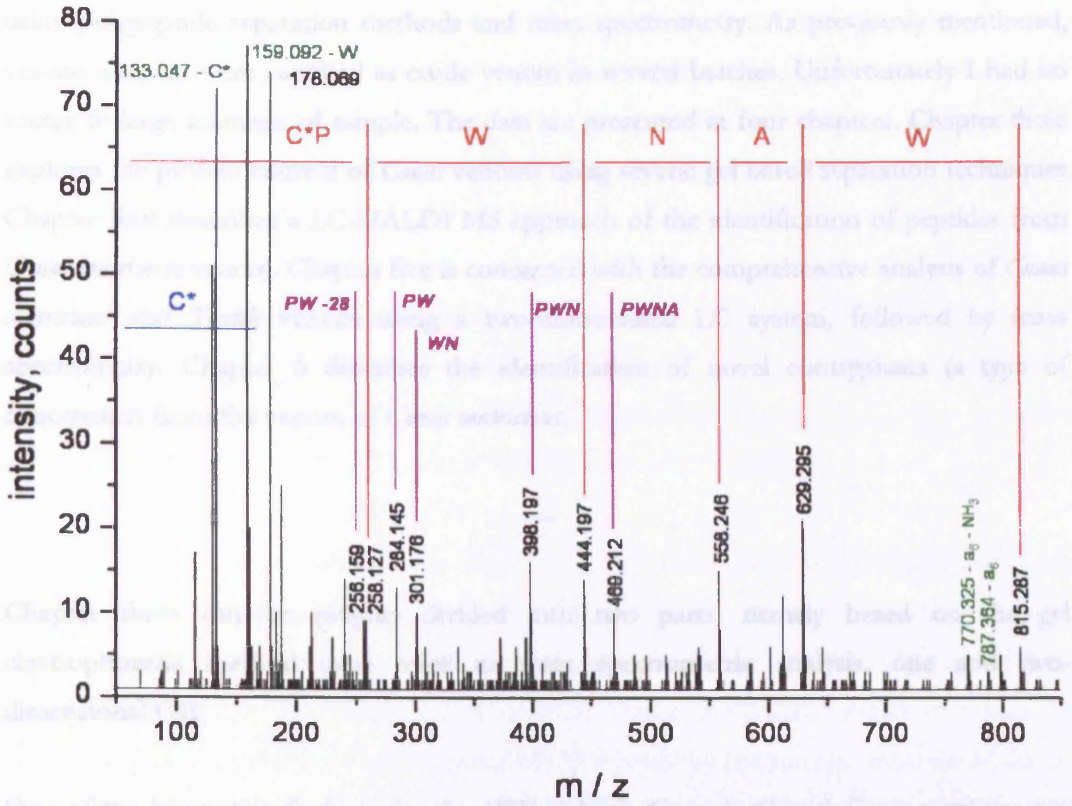


Figure 6.2.7 ESI tandem mass spectrum of contryphan with m/z 496.6001.

This figure shows the MS/MS spectrum of m/z 992.1922. The b-ions series is shown in red, the y-ion in blue. The C-terminus is amidated. Internal fragment ions with their sequence are shown in purple.

7 Concluding remarks and Outlook

The purpose of the studies presented in this thesis was the exploration of *Conus* venom using polypeptide separation methods and mass spectrometry. As previously mentioned, venom samples were supplied as crude venom in several batches. Unfortunately I had no access to large amounts of sample. The data are presented in four chapters. Chapter three explores the protein content of *Conus* venoms using several gel based separation techniques. Chapter four describes a LC-MALDI MS approach of the identification of peptides from *Conus ventricosus* venom. Chapter five is concerned with the comprehensive analysis of *Conus ventricosus* and *Textile* venom using a two-dimensional LC system, followed by mass spectrometry. Chapter 6 describes the identification of novel contryphans (a type of conotoxins) from the venom of *Conus ventricosus.*

Chapter three can be roughly divided into two parts, namely based on the gel electrophoresis method used prior to mass spectrometric analysis, one and two-dimensional GE.

One of the interesting findings for the 1DE in both *Conus textile* and *Conus ventricosus* was Vitamin K-dependent gamma-glutamyl. Conotoxins were known to have gamma-carboxyglutamate residues (Fainzilber et al., 1991). The gamma-gutamylcarboxylase responsible for this post-translational modification was found to be strictly dependent on vitamin K (Stanley et al., 1997). This carboxylase was originally found in *Conus textile* but was never reported for *Conus ventricosus* venom. Vitamin K-dependent gamma-glutamyl carboxylase has been predicted by conceptual translation in the following *Conus* species: o*maria, imperalis* and *episcopatus.* Therefore it is likely to assume that proteins like Vitamin K-dependent gamma-glutamyl carboxylase are well preserved in *Conus* species.

Among other interesting proteins identified was Protein Disulfide Isomerase (PDI), a chaperone known to facilitate the correct formation of disulfide bridges, obviously extremely important in the disulfide rich conotoxins. Further identified was Peptidyl-Prolyl cis-trans isomerase B. Conotxins are known to include L- D- conformational change of amino acids as post translational modifications.

The various methods used for identifying proteins in *Conus* venom yielded mainly so called housekeeping proteins. Actin was the most abundant protein found throughout all the tandem MS runs. For instance in case of *Conus textile* venom a total number of 26 peptides were sequenced resulting in 44.7% sequence coverage for actin with accession number 3182894. Other household proteins that were identified are vimentins. Vimentins belong to the group of intermediate filaments to which also proteins such as keratin, lamins and neurofilaments belong. Vimentins provide mechanical strength to muscle (and other) cells (Evans, 1998).

Amongst the proteins identified were also a number of inhibitors of biological processes. Examples were Protein kinase C inhibitor and trypsin inhibitor. Besides neurotoxic components, venoms from snakes, bees and spiders are known to contain protease inhibitors (Shkenderov, 1973; Zhou et al., 2004).

The overall identification rate for the 2DE analysis of the proteins in the venoms of *Conus arenatus, textile* and *nussatella* was not very high. From the 56 spots picked for in-gel digestion and subsequent MS analysis, only 12 spots resulted in an identification of four unique proteins. One interesting find is Arginine kinase (AK). AK belongs to a class of kinases that play a role in the maintenance of ATP levels by the phosphorylation of the so called "phosphagens" which then serve as a high energy source from which ATP can be rapidly replenished. In a wide variety of invertebrates, phosphoarginine is the most important phosphagen, whereas in vertebrates this is phosphocreatine. Some AKs contain an alpha-actinin like actin binding region (Pereira et al., 2000), and it has been shown that AK from mollusc muscle (Scallop, *Pecten maximus*) binds actin with reasonable affinity (Reddy et al., 1992). This protein was also identified in the 1D experiment with *Conus textile.*

Some of the proteins identified from the *Conus* venoms are normally found in the cytosol. It was therefore very likely that during the extraction procedure of the venom, cells from the actual venom duct were damaged and that their cytosolic components leaked into the venom. It is therefore difficult to establish whether the proteins identified originate from the venom. One possibility to possibly prevent the rupture of cells is to milk the snails rather than dissect the venom duct. In this process a snail is forced to inject its venom into a tube. Unfortunately we were dependent on the samples supplied.

The experiments described in chapter four were an initial attempt to analyse *Conus* venom using 2D-LC and mass spectrometry. The underlying thought of this experiment was to identify conotoxin peptides by digestion from size exclusion fractions. It is postulated in many previous papers, that *Conus* venom contains many conotoxins, up to 200 peptides were predicted. However, the actual identification, by database searching or de novo sequencing of putative conotoxins was very low. But it was a very important experience in sample handling. An added disadvantage is that manual de novo sequencing proved to be more difficult for TOF/TOF generated fragmentation data, compared to electrospray CID. TOF/TOF type fragmentation results in more internal fragments. One could argue that therefore this is also less sensitive as the sample is 'diluted'. Of course in the future this data set needs to be analysed by an adequate automatic de novo sequencing program. However batch searches performed on this data set were consistent with the findings of Chapter three. Many of the peptides originated from housekeeping proteins, primarily Actin.

An implication however of this experiment was the discovery of alkylated histidine. Alkylation of histidine has been sporadically reported in the early 60s (Fruchter and Crestfield, 1967; Szewczuk and Connell, 1965). Iodoacetamide is one of the established alkylating reagent in proteomic experiments. However, in some cases this modifies other amino acids besides the free sulfhydril groups of cysteine. This has a significant implication on database searching. Therefore, as a result of this pilot experiment, this user modification was added to some of the searches in protein prospector. However no other modified histidines were found. It was out of the scope of this work to further characterize this specific modification.

From the experiments described in Chapter five a novel conotoxin was described with the following sequence: **qCCDPCAYPCRFDYSCCY**

This conotoxin has not been reported before. The N-terminal amino acid is a pyroglutamic acid. This is a post-translational modification that occurs naturally but also quite often is user induced. However, the intact mass of the native peptide indicated a pyroglutamic acid.

The cysteine scaffold was conform with that of the M superfamily (framework III) of conotoxins (Terlau and Olivera, 2004). This framework is represented by:

– CC – C – C – CC –

To date 38 conotoxins have been reported that are classified in framework III. Amongst these toxins are for instance the μ-conotoxins which target to muscle voltage-sensitive sodium channels. The disulfide bonds between the conserved cysteines uphold the structural framework of surface loops of these toxins. However, the inter loop amino acids are hypervariable (Conticello et al., 2000; Woodward et al., 1990). Duda and coworkers reported an accelerated evolution strategy for *Conus* snail to assemble a combinatorial type library, maximising the effect to capture and kill prey (Duda and Palumbi, 1999). This work was later repeated with a larger dataset by Conticello *et al.* (Conticello et al., 2001).

One other interesting find is the confirmation of a putative conotoxin sequence found by Conticello *et al.* (Conticello et al., 2001) : **CCSWDVCDHPSCTCCG.**

In that study an EST-generated assemblage of 170 distinct conopeptides from five different *Conus* species was described. The EST library was translated from the cDNA library created from isolated mRNA from venom ducts from the different species. The preparation of *Conus Textile* venom duct cDNA library was described by Sasaki *et al.* (Sasaki et al., 1999). The mRNA needed for the creation of this library was prepared from venom ducts of 20-30 specimens each. The putative conotoxin sequences could be categorised into three cysteine frameworks, namely scaffold VI/VII, scaffold IX and scaffold III/IV (Terlau and Olivera, 2004).

In both studies it was reported that the mature toxin contained two gamma-carboxy-glutamic acid (Gla) sites at positions 68 and 73 of the precursor, however no evidence was found of this post-translational modification for our sample. Carboxy-glutamate adds 43.9898 Da to glutamic acid (E et al.). Peak lists from both electrospray and MALDI datasets were screened for masses that corresponded to one and two Gla residues (+/- 50 ppm). Gamma-carboxy glutamic acid is a post-translational modification, mainly found on proteins and peptides associated with blood coagulation or bone metabolism (Stanley et al., 1997). Gamma-carboxyglutamyl residues are good chelators of calcium ions (Sunnerhagen

et al., 1996). Blood coagulation is a process that changes circulating circumstances within the blood into an insoluble gel. The gel plugs leak in blood vessels and stops the loss of blood. This process requires coagulation factors, calcium and phospholipids. By chelating the calcium from the blood, the coagulation process is disrupted, maximising the envenomination of the prey.

The experiments described in Chapter six permitted the identification of two novel contryphans, namely:

G(I/L)COWKPW^{Br}C-NH$_2$

G(I/L)CPWKPW^{Br}C-NH$_2$

Neither Contryphan found has been reported before. Contryphans, single disulfide linked conopeptides, are know to be heavily modified. This particular peptide, a so called bromo-Contryphan was a good example. Besides one disulfide, it contained a brominated tryptophan, a C-terminal amidation and an hydroxyproline. Bromination of tryptophan is a known post-translational modification for these types of peptides. Surprisingly the same peptide revealed a differential modification of hydroxyl proline. It is of course likely that this is a 'side product' of the incomplete PTM formation.

Two more contryphans were identified, namely:

CDGPWKPWC-NH$_2$

CPWNAWC-NH$_2$

The contryphans found in this *Conus ventricosus* sample were not detected in the venom batch that was analysed in chapters four and five. The different venom batches that were analysed troughout the experiments described here were taken from different locations and at different time points. It has been postulated that there is a huge variation between different *Conus* populations. Also, the toxin expression in venom might be dependent on season and feeding. It is known that snails go without food for months. It is therefore plausible that a snail does not produce high amounts of toxins after feeding.

In conclusion, the work presented in this thesis has revealed several new conotoxins / contryphans and verified some predicted conotoxin sequences. However, the expected multitude of conotoxins that supposedly are present in *Conus* venom were not detected. The proteome of an organism, in this case *Conus*, is defined as the complete set of proteins (peptides) that can be expressed by the genetic material. Based on the number of conotoxin sequences found in the extensive separation and analysis shown by this work, the number of expressed conotoxins might well be lower than the postulated numbers in the literature.

The novel conotoxin qCCDPCAYPCRFDYSCCY that was found in the venom of *Conus ventricosus* is of course a candidate to be tested for biological activity.

To further improve the chance of finding conotoxins in complex venoms, a number of parameters could be improved. On the sample preparation side by using the milked venom rather then dissected venom ducts, to prevent the 'contamination' of the venom with materials other than that of the content of venom ducts. However in the case of *Conus ventricosus* this would be rather difficult, as these specimens only grow a few centimeters large and the amount of sample would become a considerate challenge. A pulldown and enrichment of cysteine containing peptides could be an option to circumvent the complexity of the venom. Although a good alternative for sample complexity reduction, size exclusion is a rather crude method, and into a certain extend is strong cation exchange. A good alternative for the first dimension separation would be to use free flow electrophoresis (FFE). With this separation technique, peptides and proteins are separated according to their iso-electric point. This will result in better resolved fractions which will increase the sensitivety of the overall method.

In order to increase analyse complex venom mixture in a high-throughput manner, automation is of the essence. However, to automatically search data, large and well annotated databased have to be in place. The alternative to database searching is de novo sequencing. Automating accurate de novo sequencing remains a challenge. However, with the advent of high accuracy mass spectrometers, particularly ICR-FT-MS, automated de novo sequencing seems to become more realistic, as with the increase of mass accuracy the number of false positives can be dramatically decreased.

8 Appendices

8.1 Appendix 1: Amino acid masses and more

Name residue	Symbols		Mass				
			Monoisotopic	Average	Immonium ion	y1 - ion	b1 - ion
Glycine	Gly	G	57.02146	57.0520	30.03439	76.03984	58.02929
Alanine	Ala	A	71.03711	71.0788	44.05004	90.05549	72.04494
Serine	Ser	S	87.03203	87.0782	60.04496	106.05041	88.03986
Proline	Pro	P	97.05276	97.1167	70.06569	116.07114	98.06059
Valine	Val	V	99.06841	99.1326	72.08134	118.08679	100.07624
Threonine	Thr	T	101.04768	101.1051	74.06061	120.06606	102.05551
Cysteine	Cys	C	103.00919	103.1448	76.02212	122.02757	104.01702
Leucine	Leu	L	113.08406	113.1595	86.09699	132.10244	114.09189
Isoleucine	Ile	I	113.08406	113.1595	86.09699	132.10244	114.09189
Asparagine	Asn	N	114.04293	114.1039	87.05586	133.06131	115.05076
Aspartic acid	Asp	D	115.02694	115.0886	88.03987	134.04532	116.03477
Glutamine	Gln	Q	128.05858	128.1308	101.07151	147.07696	129.06641
Lysine	Lys	K	128.09496	128.1742	101.10789	147.11334	129.10279
Glutamic acid	Glu	E	129.04259	129.1155	102.05552	148.06097	130.05042
Methionine	Met	M	131.04049	131.1986	104.05342	150.05887	132.04832
Histidine	His	H	137.05891	137.1412	110.07184	156.07729	138.06674
Phenylalanine	Phe	F	147.06841	147.1766	120.08134	166.08679	148.07624
Arginine	Arg	R	156.10111	156.1876	129.11404	175.11949	157.10894
Tyrosine	Tyr	Y	163.06333	163.1760	136.07626	182.08171	164.07116
Tryptophan	Trp	W	186.07931	186.2133	159.09224	205.09769	187.08714
Hydroxy Proline	Hyp	O	113.04770	113.1168	86.06063	132.06608	114.05553
Bromo - Tryptophan	---	---	263.98980	265.1103	237.00273	283.00818	264.99763

The molecular mass of a normally terminated and unmodified peptide or protein may be calculated by summing the masses of the appropriate amino acid residues from the table and adding the masses of H (1.00782 for monoisotopic, 1.0079 for average mass) and OH (17.00274 for monoisotopic and 17.0073 for average mass) for the N- and C-termini respectively. Also the symbols and masses of two post-translationally modified amino acids, relevant to this thesis, are given.

8.2 Appendix 2: Search results Conus ventricosus in-gel

Accession Number	Protein Description	Mass (Da)	Sequence Coverage	Mascot Protein Score	Number of Peptides	pI
3182894	Actin	41683	44.7	564.0	26	5.30
39591775	Hypothetical protein CBG18256 [Caenorhabditis briggsae]	41889	39.5	534.9	23	5.43
33346945	ubiquitin/actin fusion protein [Gymnochlora stellata]	49352	28.1	405.9	20	5.03
8895765	alpha actin [Ictalurus punctatus]	41986	33.4	364.4	24	5.30
7546413	Chain A, The Yeast Actin Val 159 Asn Mutant Complex With Human Gelsolin Segment 1	41678	16.5	286.6	9	5.44
14318479	Structural protein involved in cell polarization, endocytosis, and other cytoskeletal functions; Ac	41663	16.5	286.6	9	5.44
294852	alpha-muscle actin	41511	16.9	266.8	10	5.47
2598562	BiP [Mus musculus]	72433	8.2	265.4	6	5.10
109893	dnaK-type molecular chaperone grp78 precursor - mouse	72377	8.2	265.4	6	5.12
31981722	heat shock 70kD protein 5 (glucose-regulated protein); glucose regulated protein, 78 kDa; heat shoc	72378	8.2	265.4	6	5.01
25742763	heat shock 70kD protein 5 [Rattus norvegicus]	72302	8.3	263.7	6	5.07
45382769	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa) [Gallus gallus]	71974	8.3	263.7	6	5.12
16507237	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa); BiP; heat shock 70kD protein 5 (gluc	72288	8.3	263.7	6	5.07
45544521	heat shock protein [Numida meleagris]	72004	8.3	263.7	6	5.12
46359618	78kDa glucose regulated protein [Crassostrea gigas]	73030	9.8	262.6	6	5.02
21064361	RE12057p [Drosophila melanogaster]	37816	22.0	244.8	8	5.36
6693629	B-actin [Pagrus major]	41813	20.0	243.1	8	5.30
17647135	CG10067-PA [Drosophila melanogaster]	41808	19.9	243.1	8	5.23
576368	Chain A, Beta-Actin-Profilin Complex	41801	19.9	243.1	8	5.21
7387473	Cyl actin [Tripneustes gratilla]	41836	19.9	243.1	8	5.29
809561	gamma-actin [Mus musculus]	40992	20.4	243.1	8	5.56
19528317	LD04994p [Drosophila melanogaster]	40232	20.8	243.1	8	5.56
28336	mutant beta-actin (beta"-actin) [Homo sapiens]	41786	20.0	243.1	8	5.22
45872608	Unknown (protein for MGC:76228) [Xenopus tropicalis]	41726	20.0	243.1	8	5.29
39794594	Bactin1 protein [Danio rerio]	41740	20.0	241.4	8	5.30
9049272	beta actin [Carassius auratus]	41752	20.0	241.4	8	5.30
3386376	cytoskeletal actin 1 [Molgula oculata]	41680	20.0	241.4	8	5.29
45361511	hypothetical protein MGC75587 [Xenopus tropicalis]	41738	20.0	241.4	8	5.30
3421457	cytoplasmic actin [Dreissena polymorpha]	41826	19.9	240.4	8	5.30
7245500	Chain A, Crystal Structure Of 1:1 Complex Between Gelsolin Segment 1 And A DictyosteliumTETRAHYMENA	41619	20.0	239.7	8	5.38
7245526	Chain A, Crystal Structure Of A DictyosteliumTETRAHYMENA CHIMERA Actin (Mutant 646: Q228kT229AA230Y	41718	20.0	239.7	8	5.39
7245498	Chain A, Crystal Structure Of Dictyostelium Caatp-Actin In Complex With Gelsolin Segment 1	40846	20.4	239.7	8	5.17
2624850	Chain A, Structure Of Bovine Beta-Actin-Profilin Complex With Actin Bound Atp Phosphates Solvent Ac	41664	20.0	239.7	8	5.15
41393660	pupal-specific flight muscle actin [Aedes aegypti]	41572	20.7	235.2	8	5.29
3913786	Luminal binding protein precursor (BiP) (78 kDa glucose-	73547	6.0	231.9	5	5.03

	regulated protein homolog) (GRP 78)					
309090	A-X actin	41667	22.1	229.7	7	5.21
2724046	beta-actin [Mustela putorius furo]	35814	26.2	229.7	7	5.65
17568549	heat shock protein (73.0 kD) (hsp-3) [Caenorhabditis elegans]	72979	8.0	224.4	5	4.95
797290	muscle actin	41629	23.9	209.8	7	5.31
407521	chaperone [Saccharomyces cerevisiae]	68503	9.6	206.5	6	5.21
6319396	heat-inducible cytosolic member of the 70 kDa heat shock protein family; Ssa3p [Saccharomyces cerev	70504	9.2	206.5	6	5.05
49868	put. beta-actin (aa 27-375) [Mus musculus]	39161	21.2	200.2	8	5.78
45198632	AFR114Wp [Eremothecium gossypii]	69498	7.0	190.9	4	5.00
31212573	ENSANGP00000022307 [Anopheles gambiae]	30160	23.1	183.7	8	5.13
3550539	K15 intermediate filament type I keratin [Ovis aries]	48740	8.8	181.1	4	4.73
27525276	putative cytoskeletal actin [Ciona intestinalis]	41726	15.7	168.9	7	5.29
17061835	stress protein HSP70-1 [Xiphophorus maculatus]	70235	7.1	163.1	4	5.38
1620388	70kD heat shock protein [Takifugu rubripes]	70033	7.0	161.4	4	5.28
1170372	HEAT SHOCK PROTEIN 70 A1	70208	5.8	143.3	4	5.55
1170376	HEAT SHOCK PROTEIN 70 A2	70194	5.8	143.3	4	5.54
33591249	heat shock protein 68 [Drosophila erecta]	68292	6.0	142.6	4	5.85
123633	HEAT SHOCK 70 KD PROTEIN 4 (HSP70)	71391	7.0	142.4	4	5.25
3287859	Heat shock protein 70 (HSP70)	70025	5.8	141.6	4	5.62
38325842	heat shock protein Hsp70a [Drosophila virilis]	70132	5.8	140.9	4	5.67
38325846	heat shock protein Hsp70c [Drosophila virilis]	70090	5.8	140.9	4	5.67
38325856	heat shock protein Hsp70d [Drosophila lummei]	69895	5.8	140.9	4	5.75
38325840	heat shock protein Hsp70g [Drosophila virilis]	70131	5.8	140.9	4	5.74
1170377	HEAT SHOCK PROTEIN 70 B2	70110	5.8	139.9	4	5.68
46391140	putative Luminal binding protein [Oryza sativa (japonica cultivar-group)]	74267	4.9	136.5	3	5.09
46095698	ARF_CRYNE ADP-RIBOSYLATION FACTOR [Ustilago maydis 521]	20593	14.9	135.1	4	6.85
38104058	hypothetical protein MG04438.4 [Magnaporthe grisea 70-15]	20931	14.8	135.1	4	6.32
29789317	keratinocyte associated protein 1; cytokeratin KRT2-6HF [Mus musculus]	59704	6.4	134.5	4	8.46
18858335	bactin1; beta-actin [Danio rerio]	41726	10.4	132.7	4	5.30
14278147	Chain A, Crystal Structure Of Caenorhabditis Elegans Mg-Atp Actin Complexed With Human Gelsolin Seg	41638	10.4	132.7	4	5.30
3036959	CsCA1 [Ciona savignyi]	41698	10.4	132.7	4	5.29
39595059	Hypothetical protein CBG17727 [Caenorhabditis briggsae]	20549	15.0	131.5	4	6.43
15419605	masticatory epithelia keratin 2p [Canis familiaris]	66782	4.3	130.1	3	8.47
3891470	Chain A, Crystal Structure Of Human Galectin-7 In Complex With Galactosamine	14935	24.4	128.2	3	7.00
4504985	galectin 7; lectin galactoside-binding soluble 7 [Homo sapiens]	15066	24.3	128.2	3	7.03
46228393	ARF1/2 like small GTpase [Cryptosporidium parvum]	21443	13.2	128.1	4	5.94
30425250	RIKEN cDNA 4732495G21 gene [Mus musculus]	41977	13.8	127.6	5	5.30
4757756	annexin A2; annexin II; annexin II (lipocortin II); calpactin I, heavy polypeptide (p36); lipocorti	38580	12.4	126.0	3	7.57
18645167	ANXA2 protein [Homo sapiens]	38552	12.4	126.0	3	7.57
309215	EndoA" cytokeratin (5" end put.); putative	53210	2.5	124.2	2	5.42
18202174	14-3-3 protein sigma (Stratifin)	27832	18.1	123.1	4	4.65
631131	epithelial cell marker protein 1 - human	27759	18.1	123.1	4	4.72
16306737	SFN protein [Homo sapiens]	24321	20.8	123.1	4	4.77
5454052	stratifin [Homo sapiens]	27757	18.1	123.1	4	4.68

9055338	stratifin; 14-3-3 protein sigma; tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation pro	27696	18.1	123.1	4	4.75
229552	albumin	66088	6.0	120.6	4	5.76
123592	Heat shock 70 kDa protein	56500	6.6	118.3	3	6.45
27503439	Arf2-prov protein [Xenopus laevis]	20626	13.3	116.8	4	6.32
30749336	Chain A, Crystal Structure Of Gga1 Gat N-Terminal Region In Complex With Arf1 Gtp Form	18918	14.5	116.8	4	5.64
1065361	Chain A, Human Adp-Ribosylation Factor 1 Complexed With Gdp, Full Length Non-Myristoylated	20553	13.3	116.8	4	6.36
40889633	Chain A, Structure Of Arf1-Gdp Bound To Sec7 Domain Complexed With Brefeldin A	18789	14.6	116.8	4	5.63
39584902	Hypothetical protein CBG09004 [Caenorhabditis briggsae]	20481	13.3	116.8	4	6.15
45360689	hypothetical protein MGC76046 [Xenopus tropicalis]	20679	13.3	116.8	4	6.43
34909214	putative ADP-ribosylation factor [Oryza sativa (japonica cultivar-group)]	42184	6.4	116.8	4	10.31
17512384	BC031593 protein [Mus musculus]	59144	5.7	116.2	3	7.10
22164776	cDNA sequence BC031593 [Mus musculus]	57517	5.8	116.2	3	7.55
29028256	ARF1-like GTP-binding protein [Gossypium hirsutum]	20650	13.3	115.1	4	7.85
6573751	F20B24.7 [Arabidopsis thaliana]	21221	12.9	115.1	4	6.92
25294174	F5O8.5 protein - Arabidopsis thaliana	21553	12.8	115.1	4	6.97
25404268	probable ADP-ribolylation factor F8A5.3 [imported] - Arabidopsis thaliana	12683	21.2	115.1	4	8.76
8778579	F28C11.12 [Arabidopsis thaliana]	28379	9.8	113.4	4	9.03
23463261	vitamin K-dependent carboxylase [*Conus textile*]	93905	3.5	113.1	3	5.37
18652793	vitamin K-dependent gamma-glutamyl carboxylase [*Conus textile*]	93858	3.5	113.1	3	5.37
27805925	heat shock 70 kDa protein 8; heat shock cognate 71 kD protein; heat shock 70kd protein 10 (HSC71) [	71195	5.7	111.4	5	5.49
5729877	heat shock 70kDa protein 8 isoform 1; heat shock cognate protein, 71-kDa; heat shock 70kd protein 1	70854	5.7	111.4	5	5.37
24234686	heat shock 70kDa protein 8 isoform 2; heat shock cognate protein, 71-kDa; heat shock 70kd protein 1	53484	7.5	111.4	5	5.62
45384370	heat shock cognate 70 [Gallus gallus]	70783	5.7	111.4	5	5.47
123647	Heat shock cognate 71 kDa protein	70761	5.7	111.4	5	5.24
45544523	heat shock protein [Numida meleagris]	70827	5.7	111.4	5	5.37
42542422	Heat shock protein 8 [Mus musculus]	70828	5.7	111.4	5	5.28
13242237	heat shock protein 8; Heat shock cognate protein 70; heat shock 70kD protein 8 [Rattus norvegicus]	70827	5.7	111.4	5	5.37
16041102	hypothetical protein [Macaca fascicularis]	62378	6.5	111.4	5	5.53
17545950	PROBABLE TRANSALDOLASE PROTEIN [Ralstonia solanacearum GMI1000]	34822	6.3	104.9	3	5.50
20903595	RIKEN cDNA 2310001L23 [Mus musculus]	62806	4.7	104.6	3	8.68
22985792	COG0176: Transaldolase [Burkholderia fungorum]	38760	5.7	103.2	3	7.64
28564918	TAL1 [Saccharomyces castellii]	36911	6.0	103.2	3	6.99
202549	iodothyronine 5' monodeiodinase	54033	4.1	101.9	4	4.87
27806501	procollagen-proline, 2-oxoglutarate 4-dioxygenase [proline 4-hydroxylase], beta polypeptide [protein disulfide isomerase; thyroid hormone] [Bos taurus]	57230	3.9	101.9	4	4.80
6981324	prolyl 4-hydroxylase, beta polypeptide; Protein disulfide isomerase (Prolyl 4-hydroxylase, beta polypeptide); protein disulfide isomerase [Rattus norvegicus]	56829	3.9	101.9	4	4.82
42415475	prolyl 4-hydroxylase, beta polypeptide; protein disulfide	57023	3.9	101.9	4	4.77

	isomerase [Mus musculus]					
19880309	protein disulfide-isomerase	56975	3.9	101.9	4	4.78
339647	thyroid hormone binding protein precursor	57069	3.9	101.9	4	4.82
21703694	cognin/prolyl-4-hydroxylase/protein disulfide isomerase [Gallus gallus]	58497	3.8	100.6	5	4.75
40741908	conserved hypothetical protein [Aspergillus nidulans FGSC A4]	20600	12.4	100.1	3	9.00
27368649	H2 [Haliotis tuberculata]	380301	0.9	98.9	3	5.37
24158633	Chain A, Cyclophilin_5 From C. Elegans	19848	4.9	98.2	2	9.22
227879	cyclophilin	18995	5.0	98.2	2	8.65
17506307	CYcloPhilin (21.9 kD) (cyp-5) [Caenorhabditis elegans]	21913	4.4	98.2	2	8.98
13592067	ribosomal protein S9 [Rattus norvegicus]	22492	13.9	94.3	4	10.48
40787691	LOC395048 protein [Xenopus tropicalis]	56745	2.6	93.8	2	4.73
28436918	P4hb protein [Xenopus laevis]	57980	2.5	93.8	2	4.81
462328	HEAT SHOCK 70 KDA PROTEIN 6 (HEAT SHOCK 70 KDA PROTEIN B')	71065	3.4	93.0	2	5.77
34419635	heat shock 70kDa protein 6 (HSP70B'); heat shock 70kD protein 6 (HSP70B') [Homo sapiens]	70984	3.4	93.0	2	5.81
7672737	hsp-70-related intracellular vitamin D binding protein [Saguinus oedipus]	71112	3.4	93.0	2	6.06
38103539	hypothetical protein MG03982.4 [Magnaporthe grisea 70-15]	39700	9.3	92.6	3	5.62
34853536	similar to peroxiredoxin 1 [Rattus norvegicus]	21861	10.6	89.4	4	8.29
19912722	platyfish HSP70-1 with S-tag [Cloning vector pSTH1-GFP]	72151	4.1	88.7	5	5.43
41393662	pupal-specific flight muscle actin [Aedes aegypti]	41556	7.4	87.2	3	5.29
6980088	inner-ear cytokeratin [Rana catesbeiana]	56584	4.5	84.3	2	5.81
33598988	constitutive heat shock protein HSC70-1 [Cyprinus carpio]	70545	4.0	82.3	3	5.19
32394421	muscle-specific heat shock protein Hsc70-1 [Cyprinus carpio]	70318	4.1	82.3	3	5.19
191765	alpha-fetoprotein	47195	3.1	82.3	1	5.47
38093961	similar to protease [Mus musculus]	38658	2.3	81.9	4	9.63
112696	14-3-3 protein zeta/delta (Protein kinase C inhibitor protein-1) (KCIP-1)	27839	13.1	81.1	4	4.77
7435027	14-3-3 zeta protein - mouse	27740	13.1	81.1	4	4.72
4507953	tyrosine 3/tryptophan 5 -monooxygenase activation protein, zeta polypeptide; protein kinase C inhib	27728	13.1	81.1	4	4.73
41054687	coated vesicle membrane protein [Danio rerio]	22915	7.0	79.7	2	5.08
1352660	Cop-coated vesicle membrane protein p24 precursor	22175	7.1	79.7	2	5.07
28277266	Rnp24-prov protein [Xenopus laevis]	22826	7.0	79.7	2	4.99
39545706	L-lactate dehydrogenase A [Macroclemys temminckii]	36662	5.4	79.4	2	6.95
440306	enhancer protein	22113	8.5	77.4	2	8.16
7963723	natural killer cell enhancement factor [Oncorhynchus mykiss]	22001	8.5	77.4	2	6.51
2914482	Chain A, Complex Of The Second Kunitz Domain Of Tissue Factor Pathway Inhibitor With Porcine Trypsi	23460	12.6	76.5	2	8.26
1942351	Chain A, Crystal Structure Of The First Active Autolysate Form Of The Porcine Alpha Trypsin	13284	22.4	76.5	2	7.83
494360	Chain A, Trypsin (E.C.3.4.21.4) Complexed With Inhibitor From Bitter Gourd	23458	12.6	76.5	2	8.60
999627	Chain B, Porcine E-Trypsin (E.C.3.4.21.4)	8814	34.1	76.5	2	6.67
3318722	Chain E, Leech-Derived Tryptase InhibitorTRYPSIN COMPLEX	23457	12.6	76.5	2	8.26
136429	Trypsin precursor	24394	12.1	76.5	2	7.00
22086550	heat shock protein Hsp90 [Achlya ambisexualis]	80120	2.7	75.8	5	5.00

22086553	heat shock protein Hsp90 [Achlya ambisexualis]	80134	2.7	75.8	5	5.01
1314736	Der f 3 mite allergen	24938	3.9	74.5	3	5.57
2507248	Mite allergen Der f 3 precursor (Der f III)	27657	3.5	74.5	3	5.36
729315	Mite allergen Der p 3 precursor (Der p III)	28042	3.4	74.5	3	8.08
14423685	Mite allergen Eur m 3 precursor	28020	3.4	74.5	3	5.80
32527715	Ac2-048 [Rattus norvegicus]	33643	6.0	73.9	2	9.72
38372905	cell line NK14 derived transforming oncogene [Mus musculus]	23653	8.7	73.9	2	9.15
5669640	ethylene-responsive small GTP-binding protein [Lycopersicon esculentum]	23755	8.3	73.9	2	8.37
23463313	GTPase Rab8b [Rattus norvegicus]	23588	8.7	73.9	2	9.15
24306110	GTP-binding protein [Pichia angusta]	22561	8.9	73.9	2	5.48
1362065	GTP-binding protein GTP11 - garden pea	23687	8.4	73.9	2	7.66
1362067	GTP-binding protein GTP13 - garden pea	23791	8.4	73.9	2	7.66
1362066	GTP-binding protein GTP6 - garden pea	23924	8.3	73.9	2	7.63
103720	GTP-binding protein o-rab1 - electric ray (Discopyge ommata)	22218	9.0	73.9	2	5.54
92339	GTP-binding protein rab1B - rat	22176	9.0	73.9	2	5.55
466171	GTP-binding protein ypt1	22462	8.9	73.9	2	5.47
32418088	GTP-BINDING PROTEIN YPT1 [Neurospora crassa]	26693	7.4	73.9	2	6.82
4886443	hypothetical protein [Homo sapiens]	13894	14.0	73.9	2	5.44
7498104	hypothetical protein D1037.4 - Caenorhabditis elegans	25664	8.0	73.9	2	5.26
46123663	hypothetical protein FG06209.1 [Gibberella zeae PH-1]	22694	8.9	73.9	2	5.95
38106736	hypothetical protein MG06135.4 [Magnaporthe grisea 70-15]	22719	8.7	73.9	2	6.60
38109427	hypothetical protein MG06962.4 [Magnaporthe grisea 70-15]	22448	8.9	73.9	2	5.27
46099538	hypothetical protein UM03865.1 [Ustilago maydis 521]	22860	8.6	73.9	2	5.69
37360618	mKIAA3012 protein [Mus musculus]	27317	7.1	73.9	2	7.68
33146687	putative ethylene-responsive small GTP-binding protein [Oryza sativa (japonica cultivar-group)]	23978	8.4	73.9	2	7.63
14475537	putative Rab/GTPase [Colletotrichum lindemuthianum]	22533	8.9	73.9	2	6.60
25150215	RAB family member (24.0 kD) (rab-8) [Caenorhabditis elegans]	24007	8.5	73.9	2	7.64
4758988	RAB1A, member RAS oncogene family; RAB1, member RAS oncogene family [Homo sapiens]	22663	8.8	73.9	2	5.93
13569962	RAB1B, member RAS oncogene family; small GTP-binding protein [Homo sapiens]	22157	9.0	73.9	2	5.55
1619841	rab1-like [Caenorhabditis elegans]	7735	26.1	73.9	2	4.57
27924279	Rab1-prov protein [Xenopus laevis]	22647	8.8	73.9	2	5.93
1370190	RAB8A [Lotus corniculatus var. japonicus]	23761	8.3	73.9	2	6.60
7706563	RAB8B, member RAS oncogene family; RAB-8b protein [Homo sapiens]	23569	8.7	73.9	2	9.15
1370194	RAB8C [Lotus corniculatus var. japonicus]	23520	8.5	73.9	2	7.64
1370196	RAB8D [Lotus corniculatus var. japonicus]	23613	8.4	73.9	2	6.74
1370198	RAB8E [Lotus corniculatus var. japonicus]	23700	8.4	73.9	2	6.74
1619851	rab8-like [Caenorhabditis elegans]	14194	14.6	73.9	2	7.74
2500076	Ras-like GTP-binding protein YPT1	22278	9.0	73.9	2	7.60
15238542	Ras-related GTP-binding family protein [Arabidopsis thaliana]	23819	8.3	73.9	2	8.35
15231847	Ras-related GTP-binding protein, putative [Arabidopsis thaliana]	23924	8.3	73.9	2	8.35
15231322	Ras-related protein (ARA-3) / small GTP-binding protein, putative [Arabidopsis thaliana]	23820	8.3	73.9	2	7.66
131785	Ras-related protein ORAB-1	22319	8.9	73.9	2	5.54
131787	Ras-related protein Rab-1A	22748	8.8	73.9	2	5.95
464524	Ras-related protein Rab-1A	22745	8.8	73.9	2	5.52

131803	Ras-related protein Rab-1B	22149	9.0	73.9	2	5.55
3024527	Ras-related protein RAB1BV	23772	8.4	73.9	2	7.66
32411557	RAS-RELATED PROTEIN RAB1BV [Neurospora crassa]	22541	8.8	73.9	2	5.58
1710002	Ras-related protein Rab-8 (Oncogene c-mel)	23542	8.7	73.9	2	9.26
21313162	RIKEN cDNA 1110011F09 [Mus musculus]	22173	9.0	73.9	2	5.55
11558649	secretion related GTPase (SrgB) [Aspergillus niger]	22289	9.0	73.9	2	5.14
27817324	SI:zC101N13.3 (novel protein similar to human RAS oncogene family member RAB1B) [Danio rerio]	22274	9.0	73.9	2	5.56
41055496	similar to RAB1, member RAS oncogene family [Danio rerio]	22370	9.0	73.9	2	5.55
27709432	similar to Ras-related protein Rab-1B [Rattus norvegicus]	22163	9.0	73.9	2	5.55
7438394	small GTP-binding protein rab-1 - turnip	25005	8.1	73.9	2	6.17
11558500	small GTP-binding protein YPT1 [Hypocrea jecorina]	22407	8.9	73.9	2	5.47
40739629	YPT1_NEUCR GTP-binding protein ypt1 [Aspergillus nidulans FGSC A4]	22204	9.0	73.9	2	5.53
46138717	YPT1_NEUCR GTP-binding protein ypt1 [Gibberella zeae PH-1]	22411	8.9	73.9	2	5.28
19115492	ypt1-related protein 2 [Schizosaccharomyces pombe]	22741	9.0	73.9	2	5.59
7533034	YptA [Aspergillus awamori]	22321	9.0	73.9	2	5.14
39579623	Hypothetical protein CBG22322 [Caenorhabditis briggsae]	41252	5.6	73.5	3	5.88
27357027	alpha tubulin [Oncorhynchus mykiss]	21399	14.6	72.8	3	5.85
340019	alpha-tubulin	38102	8.2	72.8	3	4.92
37779014	Cra10 alpha tubulin [Pagrus major]	24124	12.9	72.8	3	5.33
22595342	serine protease [Dermatophagoides pteronyssinus]	26311	3.7	72.8	3	6.82
46016010	Chain C, Crystal Structure Analysis Of The Site Specific Mutant (Q253c) Of Bovine Carbonic Anhydras	29199	7.3	72.5	2	6.63
6006601	beta-mannanase [Thermotoga neapolitana]	91483	1.6	72.4	3	6.17
2429092	beta-xylosidase [Thermotoga neapolitana]	86819	1.7	72.4	3	5.73
39588485	Hypothetical protein CBG01077 [Caenorhabditis briggsae]	22561	8.8	72.2	2	5.52
39586275	Hypothetical protein CBG12025 [Caenorhabditis briggsae]	23980	8.5	72.2	2	7.64
16933567	mel transforming oncogene; mel transforming oncogene (derived from cell line NK14)- RAB8 homolog; r	23653	8.7	72.2	2	9.15
17558550	RAB family member (22.5 kD) (rab-1) [Caenorhabditis elegans]	22531	8.8	72.2	2	5.52
234746	RAS-related protein MEL [Homo sapiens]	23582	8.7	72.2	2	9.35
18447913	ras-related protein RAB8-1 [Nicotiana tabacum]	23959	8.3	72.2	2	7.66
15077428	small GTP-binding protein Ypt1p [Candida albicans]	23005	8.7	72.2	2	5.27
11277107	heat-shock protein 70 [imported] - Petrosia ficiformis (fragment)	51549	4.7	72.1	3	5.95
28277305	Vim4 protein [Xenopus laevis]	53507	4.1	71.9	3	5.12
138532	Vimentin 4	53464	4.1	71.9	3	5.08
138531	Vimentins 1 and 2	52812	4.1	71.9	3	5.16
33591156	thioredoxin peroxidase [Ixodes ricinus]	19134	10.7	71.3	4	6.93
27735456	Alphatub84b-prov protein [Xenopus laevis]	50147	6.2	71.1	3	4.94
15988311	Chain A, Refined Structure Of Alpha-Beta Tubulin From Zinc-Induced Sheets Stabilized With Taxol	50022	6.2	71.1	3	4.91
3745821	Chain A, Tubulin Alpha-Beta Dimer, Electron Diffraction	48785	6.4	71.1	3	5.18
68288	carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)	28963	7.3	70.8	2	7.93
30466252	carbonic anhydrase II [Bos taurus]	29096	7.3	70.8	2	6.41
46016008	Chain A, Crystal Structure Analysis Of Bovine Carbonic Anhydrase Ii	28965	7.3	70.8	2	6.40
12214171	putative small GTP-binding protein (rab1b) [Homo sapiens]	18648	11.0	70.5	2	9.87

13537192	SCCA1b [Homo sapiens]	38495	5.9	70.0	2	6.28
5902072	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 3; squamous cell carcinoma a	44537	5.1	70.0	2	6.35
897844	squamous cell carcinoma antigen	13407	17.1	70.0	2	9.16
10121745	AAG13352.1 | skeletal alpha-actin [Gillichthys mirabilis]	13956	19.4	69.6	4	8.49
42525543	glycosyl hydrolase, family 3 [Treponema denticola ATCC 35405]	61795	2.2	69.6	3	6.59
44829400	peroxiredoxin 1 [Ixodes ricinus]	18977	10.6	69.6	4	11.80
2143554	14-3-3 protein isoform zeta - rat (fragment)	21471	15.4	69.0	3	4.67
42523784	thiosulfate sulfurtransferase [Bdellovibrio bacteriovorus HD100]	31343	4.3	68.6	1	9.45
2118384	leupin precursor - human	44873	5.1	68.4	2	5.86
13537194	SCCA2b [Homo sapiens]	42260	5.4	68.4	2	5.74
131774	20 kDa protein in rubredoxin operon (ORF C)	20023	8.4	68.1	3	4.96
22208480	putative stripe rust resistance protein Yr10 [Sorghum bicolor]	99421	3.7	68.0	5	8.68
44917135	14-3-3 a-1 protein [Nicotiana tabacum]	28747	9.4	67.8	3	4.70
44917143	14-3-3 c-2 protein [Nicotiana tabacum]	29362	9.2	67.8	3	4.78
44917151	14-3-3 e-1 protein [Nicotiana tabacum]	29365	9.2	67.8	3	4.72
44917153	14-3-3 e-2 protein [Nicotiana tabacum]	29337	9.2	67.8	3	4.72
44917155	14-3-3 f-1 protein [Nicotiana tabacum]	29078	9.3	67.8	3	4.76
45476403	14-3-3 f-2 protein [Nicotiana tabacum]	27419	9.9	67.8	3	4.80
44917163	14-3-3 i-1 protein [Nicotiana tabacum]	29368	9.2	67.8	3	4.79
44917165	14-3-3 i-2 protein [Nicotiana tabacum]	29454	9.2	67.8	3	4.77
7435015	14-3-3 protein - barley	29924	9.1	67.8	3	4.73
26454607	14-3-3 protein 2	28861	9.4	67.8	3	4.72
3041662	14-3-3 protein 3 (PBLT3)	29287	9.2	67.8	3	4.74
1168191	14-3-3 protein 4 (PBLT4)	29281	9.2	67.8	3	4.69
3023182	14-3-3 protein 5	28736	9.4	67.8	3	4.68
26454608	14-3-3 protein 6	28949	9.3	67.8	3	4.70
2895518	14-3-3 protein epsilon [Xenopus laevis]	29168	9.4	67.8	3	4.68
18396217	14-3-3 protein GF14 nu (GRF7) [Arabidopsis thaliana]	29806	9.1	67.8	3	4.74
466334	14-3-3 protein homologue	25776	10.5	67.8	3	4.88
41056815	14-3-3 protein isoform 20R [Solanum tuberosum]	28748	9.4	67.8	3	4.72
7435010	14-3-3 protein tft2 - tomato	28804	9.4	67.8	3	4.63
6321025	14-3-3 protein, major isoform; binds proteins and DNA, involved in regulation of many processes inc	30073	9.0	67.8	3	4.82
46096636	1433_CANAL 14-3-3 protein homolog [Ustilago maydis 521]	29263	9.2	67.8	3	4.76
40741654	1433_TRIHA 14-3-3 PROTEIN HOMOLOG (TH1433) [Aspergillus nidulans FGSC A4]	29098	9.2	67.8	3	4.75
34452069	14-3-3E1 protein [Oncorhynchus mykiss]	29298	9.3	67.8	3	4.67
34452071	14-3-3E2 protein [Oncorhynchus mykiss]	28988	9.4	67.8	3	4.70
1702985	14-3-3-like protein	28716	9.5	67.8	3	4.82
3023190	14-3-3-LIKE PROTEIN 16R	28918	9.3	67.8	3	4.74
38569374	14-3-3-like protein 2 [Paracoccidioides brasiliensis]	29624	9.1	67.8	3	4.68
3023194	14-3-3-LIKE PROTEIN A (SGF14A)	29031	9.3	67.8	3	4.72
1168189	14-3-3-LIKE PROTEIN A (VFA-1433A)	29402	9.2	67.8	3	4.71
3912948	14-3-3-LIKE PROTEIN B	28809	9.4	67.8	3	4.70
2492487	14-3-3-LIKE PROTEIN B (14-3-3B)	29673	9.2	67.8	3	4.67
3023189	14-3-3-like protein C (14-3-3-like protein B)	29346	9.2	67.8	3	4.78
3912950	14-3-3-LIKE PROTEIN E	30551	8.8	67.8	3	4.98

3912951	14-3-3-LIKE PROTEIN F	29207	9.3	67.8	3	4.76
1702986	14-3-3-like protein GF14 chi (General regulatory factor 1)	29883	9.0	67.8	3	4.68
18413181	14-3-3-like protein GF14 chi / general regulatory factor 1 (GRF1) [Arabidopsis thaliana]	29913	9.0	67.8	3	4.68
1168200	14-3-3-like protein GF14 psi (General regulatory factor 3) (14-3-3-like protein RCI1)	28589	9.4	67.8	3	4.68
1345587	14-3-3-LIKE PROTEIN GF14-6	29644	9.2	67.8	3	4.76
2492489	14-3-3-LIKE PROTEIN RA215	28577	9.4	67.8	3	4.78
543711	14-3-3-LIKE PROTEIN S94	29112	9.2	67.8	3	4.77
45201203	AGR107Cp [Eremothecium gossypii]	28389	9.6	67.8	3	4.87
13430385	ARTA [Emericella nidulans]	29097	9.2	67.8	3	4.83
14532442	At1g35160/T32G9_30 [Arabidopsis thaliana]	30160	9.0	67.8	3	4.79
671634	BMH1 [Saccharomyces cerevisiae]	30158	9.0	67.8	3	4.87
683696	BMH2 [Saccharomyces cerevisiae]	31148	8.8	67.8	3	4.82
13447104	GF14 omega [Brassica napus]	29108	9.2	67.8	3	4.68
38110176	hypothetical protein MG01588.4 [Magnaporthe grisea 70-15]	30075	8.8	67.8	3	4.93
13702816	putative 14-3-3 protein [Oryza sativa]	29160	9.2	67.8	3	4.81
28279518	Similar to tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypept	29054	9.4	67.8	3	4.65
9798603	TaWIN1 [Triticum aestivum]	29377	9.0	67.8	3	4.75
9798605	TaWIN2 [Triticum aestivum]	28680	9.3	67.8	3	4.80
6688554	tft3 14-3-3 protein [Lycopersicon esculentum]	25977	10.4	67.8	3	4.94
29367361	WIN2-like protein [Oryza sativa (japonica cultivar-group)]	28982	9.1	67.8	3	4.85
23613323	hypothetical protein [Plasmodium falciparum 3D7]	214645	1.2	67.6	3	8.65
27497749	elongation factor-1 alpha [Chrysolina coerulans]	15539	14.3	66.4	2	7.77
15986653	14-3-3 epsilon 2 [Schistosoma mansoni]	28414	9.6	66.1	3	5.20
530049	14-3-3 protein	26279	10.4	66.1	3	4.80
112684	14-3-3-LIKE PROTEIN A (14-3-3A)	29334	9.2	66.1	3	4.83
11262432	DNA damage checkpoint protein rad24 - fission yeast (Schizosaccharomyces pombe)	30357	8.9	66.1	3	4.84
19115079	dna damage checkpoint protein Rad24p [Schizosaccharomyces pombe]	30064	8.9	66.1	3	4.66
17542208	endoplasmin (87.1 kD) (4L887) [Caenorhabditis elegans]	87058	2.4	65.8	5	4.97
39593714	Hypothetical protein CBG06014 [Caenorhabditis briggsae]	87138	2.4	65.8	5	5.03
16596797	hemocyanin [Octopus dofleini]	334668	0.7	65.5	2	5.61
6685487	Hemocyanin G-type, units Oda to Odg	331711	0.7	65.5	2	5.58
32266201	Inosinic acid dehydrogenase GuaB [Helicobacter hepaticus ATCC 51449]	52001	2.9	65.3	2	6.77
38077190	expressed sequence AI507495 [Mus musculus]	56715	3.7	65.1	2	7.08
6492278	coatomer complex COPI delta-COP subunit [Drosophila melanogaster]	54198	2.4	65.0	1	5.56
3256111	EG:63B12.10 [Drosophila melanogaster]	57782	2.3	65.0	1	5.85
38146975	DNA polymerase I [Caldicellulosiruptor saccharolyticus]	97778	1.9	64.9	4	5.99
23619253	helicase, putative [Plasmodium falciparum 3D7]	136547	1.0	64.8	3	8.21
33872678	YWHAZ protein [Homo sapiens]	31711	7.9	64.5	3	4.94
123615	Heat shock 70 kDa protein, mitochondrial precursor	71102	2.9	64.1	2	7.49
25553516	mitochondrial HSP70 [Trypanosoma congolense]	71625	2.9	64.1	2	5.70
38104826	hypothetical protein MG09353.4 [Magnaporthe grisea 70-15]	77045	1.4	63.7	3	5.99
19114004	atp-dependent dna helicase hus2 [Schizosaccharomyces pombe]	149553	0.7	63.7	2	6.73
17561130	emp24/gp25L/p24 family (23.7 kD) (5L804) [Caenorhabditis elegans]	23665	7.4	63.3	2	7.06

39592002	Hypothetical protein CBG23171 [Caenorhabditis briggsae]	23618	7.4	63.3	2	7.02
3087752	S31 protein [Cyprinus carpio]	14867	11.7	63.3	2	4.91
33241320	Molecular chaperone, DnaK [Prochlorococcus marinus subsp. marinus str. CCMP1375]	68186	2.7	63.2	1	4.78
9909982	putative XIRG protein [Xenopus laevis]	54675	3.2	63.2	2	6.34
17433174	Transaldolase	37452	4.2	63.1	2	6.57
236793	13 kDa calgizzarin-like protein [rabbits, lung, Peptide Partial, 35 aa]	3923	45.7	62.9	1	4.83
115499	Calgizzarin (S100C protein)	11422	15.7	62.9	1	6.73
34811429	Chain A, Solution Structure Of Rabbit Apo-S100a11 (19 Models)	11291	15.8	62.9	1	6.69
7388240	Putative S100 calcium-binding protein A11 pseudogene	11280	15.7	62.9	1	7.77
20874201	S100 calcium binding protein A11 (calizzarin) [Mus musculus]	11075	16.3	62.9	1	5.28
25023967	similar to endothelial monocyte-activating polypeptide [Mus musculus]	11002	16.3	62.9	1	5.73
19113132	cell division control protein 7 [Schizosaccharomyces pombe]	119216	1.6	62.5	5	7.54
86248	dolichyl-diphosphooligosaccharide-protein glycotransferase (EC 2.4.1.119) glycosylation site-binding chain precursor - chicken	56929	2.6	62.3	2	4.84
45382295	glycosylation site-binding protein [Gallus gallus]	56856	2.6	62.3	2	4.84
27370850	Hspa5-prov protein [Xenopus laevis]	72375	2.7	62.1	2	5.03
860916	60C beta tubulin [Drosophila melanogaster]	48805	4.4	62.0	2	4.69
30088884	beta tubulin [Aplysia californica]	50286	4.2	62.0	2	4.73
1066143	beta-tublin [Haliotis discus]	38269	5.6	62.0	2	5.85
27227551	class II beta tubulin isotype [Homo sapiens]	49884	4.3	62.0	2	4.82
4884102	hypothetical protein [Homo sapiens]	36792	5.8	62.0	2	4.85
45361395	hypothetical protein MGC75628 [Xenopus tropicalis]	49889	4.3	62.0	2	4.78
13097483	Tubb2 protein [Mus musculus]	33979	6.4	62.0	2	4.81
4105827	nuclear alpha-tubulin [Guillardia theta]	42355	4.2	61.8	2	6.14
13879585	1110014C03Rik protein [Mus musculus]	11327	16.0	61.7	2	7.82
16758214	integral membrane protein Tmp21-I (p23) [Rattus norvegicus]	24842	6.8	61.7	2	6.02
3915137	Transmembrane protein Tmp21 precursor (21 kDa Transmembrane trafficking protein)	23262	7.4	61.7	2	6.03
40555903	Transmembrane trafficking protein [Mus musculus]	24955	6.8	61.7	2	6.25
37748485	Wu:fe06g04 protein [Danio rerio]	24426	7.2	61.7	2	6.37
39595426	Hypothetical protein CBG04075 [Caenorhabditis briggsae]	49948	3.1	61.6	2	5.08
39596178	Hypothetical protein CBG16131 [Caenorhabditis briggsae]	49947	3.1	61.6	2	5.01
38109071	hypothetical protein MG06650.4 [Magnaporthe grisea 70-15]	51390	3.0	61.6	2	5.33
46107512	TBA_SORMA Tubulin alpha chain [Gibberella zeae PH-1]	49939	3.1	61.6	2	4.98
40746566	TBA1_EMENI TUBULIN ALPHA-1 CHAIN [Aspergillus nidulans FGSC A4]	50327	3.1	61.6	2	4.92
28261777	ycf1 protein [Atropa belladonna]	223728	0.8	61.5	3	9.70
16945685	disulfide isomerase [Ostertagia ostertagi]	54972	2.8	61.5	3	4.93
18086539	reverse transcriptase [Human immunodeficiency virus type 1]	26147	7.9	61.0	4	9.32
6689664	p24 delta1 putative cargo receptor [Xenopus laevis]	23930	7.2	60.7	2	6.92
46318644	Enolase [Burkholderia cepacia R1808]	45642	2.6	60.5	2	4.77
23612644	integral membrane protein, putative [Plasmodium falciparum 3D7]	224933	0.9	60.3	3	8.74
23491117	235 kDa rhoptry protein [Plasmodium yoelii yoelii]	264686	1.1	60.3	6	6.35
2981970	Chain A, An Avian Class-Mu Glutathione S-Transferase, Cgstm1-1 At 1.94 Angstrom Resolution	25745	4.1	60.1	1	7.00

10120620	Chain A, Tyr115, Gln165 And Trp209 Contribute To The 1,2-Epoxy-3-(P- Nitrophenoxy)propane Conjugating Activities Of Glutathione S-Transferase Cgstm1-1	25730	4.1	60.1	1	7.00
5822511	Chain B, Ligand-Free Heterodimeric Human Glutathione S-Transferase M2-3 (Ec 2.5.1.18), Monoclinic C	26411	4.0	60.1	1	5.37
8489487	class mu glutathione S-transferase [Capra hircus]	22332	4.8	60.1	1	8.57
255725	glutathione S-transferase, GST=isoenzyme M4 [Syrian golden hamsters, Peptide Partial, 21 aa, segmen	2439	42.9	60.1	1	10.45
204499	glutathione S-transferase Y-b subunit (EC 2.5.1.18)	21871	4.8	60.1	1	7.82
423912	mu-class glutathione S-transferase hGSTYBX - hamster	25551	4.1	60.1	1	6.90
28494710	RIKEN cDNA 0610005A07 [Mus musculus]	25693	4.1	60.1	1	6.34
15897104	Hypothetical protein [Sulfolobus solfataricus]	69157	2.7	60.1	5	8.62
13473927	DNA-directed DNA polymerase [Mesorhizobium loti MAFF303099]	106058	1.8	59.9	3	6.80
18311435	probable lysine decarboxylase [Clostridium perfringens]	54681	3.2	59.7	4	5.93
347527	ribosomal protein S3	10679	18.9	59.6	2	9.69
231660	Hypothetical 226 kDa protein ycf1 (ORF 1901)	225999	0.6	59.4	2	9.74
46121903	hypothetical protein FG05329.1 [Gibberella zeae PH-1]	111987	1.2	58.9	2	5.68
18312252	conserved hypothetical protein [Pyrobaculum aerophilum]	49023	2.1	58.9	2	5.04
23488517	hypothetical protein [Plasmodium yoelii yoelii]	457509	0.5	58.7	6	5.36
45190352	AEL255Wp [Eremothecium gossypii]	21400	8.5	58.6	2	10.27
4557966	Chain A, Ligand-Free Human Glutathione S-Transferase M2-2 (E.C.2.5.1.18), Monoclinic Crystal Form	25597	4.1	58.4	1	6.02
21743012	OSJNBb0013O03.2 [Oryza sativa (japonica cultivar-group)]	132160	1.2	58.4	3	8.86
1752728	alpha-aminoadipyl-cysteinyl-valine synthetase [Lysobacter lactamgenus]	411361	0.5	58.2	4	5.47
15922299	422aa long hypothetical interferon-gamma inducible protein [Sulfolobus tokodaii]	49426	3.3	58.1	3	6.66
45184912	AAR089Cp [Eremothecium gossypii]	186517	1.1	58.1	5	7.54
19698783	thiredoxin peroxidase [Acanthocheilonema viteae]	27703	6.1	58.1	2	5.10
17565948	predicted CDS, putative protein family member (5S783) [Caenorhabditis elegans]	74114	1.4	58.0	3	4.61
19111953	putative RNA binding protein [Schizosaccharomyces pombe]	31418	4.4	57.9	3	7.27
42519216	signal peptidase I [Lactobacillus johnsonii NCC 533]	21926	5.3	57.7	3	9.74
46323625	COG1012: NAD-dependent aldehyde dehydrogenases [Burkholderia cepacia R1808]	50608	2.7	57.6	2	5.74
32484222	Gstm2-prov protein [Xenopus laevis]	25375	4.1	57.4	1	7.64
46097256	hypothetical protein UM01791.1 [Ustilago maydis 521]	122733	0.9	57.4	2	7.92
12733945	40S ribosomal protein S15a [Platichthys flesus]	9145	21.3	57.3	2	10.43
36142	ribosomal protein homologous to yeast S24 [Homo sapiens]	14707	13.2	57.3	2	10.23
18000277	ribosomal protein S15 isoform [Lapemis hardwickii]	14846	13.1	57.3	2	10.07
14165469	ribosomal protein S15a; 40S ribosomal protein S15a [Homo sapiens]	14830	13.1	57.3	2	10.14
30109302	Rps15a protein [Mus musculus]	15610	12.3	57.3	2	10.01
45534427	hypothetical protein Mflo061301 [Mesoplasma florum L1]	68947	2.4	56.9	4	5.40
46129441	RS9_PODAN 40S ribosomal protein S9 (S7) [Gibberella zeae	21947	8.4	56.9	2	10.5

	PH-1]					3
14701840	MB2 [Plasmodium falciparum]	106077	2.2	56.7	5	6.85
23473347	Signal transduction histidine kinase [Desulfovibrio desulfuricans G20]	51082	3.3	56.6	2	5.76
38105537	hypothetical protein MG03548.4 [Magnaporthe grisea 70-15]	189263	0.9	56.5	4	5.51
30313543	mitochondrial malate dehydrogenase precursor [Plicopurpura patula]	24405	6.1	56.5	2	6.84
11493210	CAC17476.1 | alpha tubulin [Ustilago maydis]	49072	5.4	56.4	3	4.92
28302252	Calr-prov protein [Xenopus laevis]	48491	4.1	56.4	2	4.42
28277246	Crc-prov protein [Xenopus laevis]	48998	4.1	56.4	2	4.41
20896331	RIKEN cDNA 4921538N17 [Mus musculus]	148522	2.1	56.3	5	6.36
11138320	vf14-3-3c protein [Vicia faba]	29664	6.8	56.2	2	4.79
15678305	bacteriochlorophyll synthase 43 kDa subunit [Methanothermobacter thermautotrophicus]	42199	5.0	56.0	2	4.75
2895520	14-3-3 protein zeta [Xenopus laevis]	27712	9.0	56.0	2	4.73
34783865	MGC64423 protein [Xenopus laevis]	27741	9.0	56.0	2	4.70
27370992	Ywhaz-prov protein [Xenopus laevis]	27740	9.0	56.0	2	4.72
32398830	rhoptry protein, possible [Cryptosporidium parvum]	161902	1.6	55.9	4	6.01
31982284	UV radiation resistance associated [Mus musculus]	77514	1.0	55.8	1	8.20
6635207	Uvrag [Mus musculus]	38868	2.1	55.8	1	8.81
6319146	H2A protein [Oryza sativa]	14485	6.5	55.7	1	10.18
15224957	histone H2A, putative [Arabidopsis thaliana]	14361	6.6	55.7	1	10.48
15232536	histone H2A.F/Z [Arabidopsis thaliana]	14532	6.6	55.7	1	10.32
37534390	putative histone H2A [Oryza sativa (japonica cultivar-group)]	14603	6.5	55.7	1	10.39
198578	ribosomal protein	17730	5.9	55.6	1	11.12
28626512	ribosomal protein S18 [Danio rerio]	17721	5.9	55.6	1	10.99
41150652	similar to ribosomal protein [Homo sapiens]	17628	5.9	55.6	1	10.34
32469491	optineurin; TFIIIA-INTP [Mus musculus]	66976	2.9	55.6	2	5.21
23598413	hemocyanin [Euprymna scolopes]	14367	8.0	55.4	2	6.57
6685493	Hemocyanin, beta-C chain unit G	46182	2.5	55.4	2	5.48
419976	calreticulin (clone 8) - African clawed frog (fragment)	45246	4.4	55.4	2	4.40
15384299	variable surface lipoprotein Vsp422-10 [Mycoplasma bovis]	29812	2.8	55.2	1	8.37
21224465	putative sensor kinase [Streptomyces coelicolor A3(2)]	41405	3.1	55.2	2	8.81
3426021	heat shock 70 kD protein cognate [Bombyx mori]	73052	2.7	55.1	2	5.13
27260894	heat shock cognate 70 protein [Spodoptera frugiperda]	73064	2.7	55.1	2	5.20
6681668	sea urchin Arp2 (SUArp2) [Hemicentrotus pulcherrimus]	44145	4.6	55.1	3	6.73
15791804	succinate dehydrogenase flavoprotein subunit [Campylobacter jejuni subsp. jejuni NCTC 11168]	66330	3.6	55.0	3	6.43
21230675	hypothetical protein [Xanthomonas campestris pv. campestris str. ATCC 33913]	31201	2.4	54.9	1	6.24
29375773	conserved hypothetical protein [Enterococcus faecalis V583]	33825	2.3	54.6	1	8.56
29346075	hypothetical protein [Bacteroides thetaiotaomicron VPI-5482]	25484	3.3	54.6	1	8.87
10338322	14-3-3 epsilon [Schistosoma japonicum]	28529	8.0	54.6	3	5.64
42560899	DNA polymerase III alpha chain [Mycoplasma mycoides subsp.	170741	0.6	54.5	2	6.10

	mycoides SC]					
17559380	acyltransferase (82.0 kD) (5F913) [Caenorhabditis elegans]	82019	1.7	54.5	3	8.53
153462	sporulation protein 1422	49522	5.7	54.3	3	6.74
28630917	ADP-ribosylation factor [Gossypium hirsutum]	5426	22.4	54.2	1	10.12
29245482	GLP_334_11456_12031 [Giardia lamblia ATCC 50803]	21751	5.8	54.2	1	8.48
28436853	LOC398551 protein [Xenopus laevis]	20350	6.2	54.2	1	6.81
4102193	Tcarf [Trypanosoma cruzi]	7402	16.7	54.2	1	9.45
23017275	Leucyl-tRNA synthetase [Thermobifida fusca]	92582	3.9	54.1	3	5.23
40745738	GR78_NEUCR 78 KDA GLUCOSE-REGULATED PROTEIN HOMOLOG PRECURSOR (GRP 78) (IMMUNOGLOBULIN HEAVY CHAIN	73652	3.3	54.1	2	4.84
38081072	similar to myosin heavy chain [Mus musculus]	270321	1.1	53.9	6	6.51
41393050	polyprotein [Soybean mosaic virus]	354371	0.6	53.9	4	8.55
15895958	IMP dehydrogenase [Clostridium acetobutylicum]	52155	2.5	53.9	3	7.05
212997	keratin	49148	4.1	53.9	3	5.60
45198425	AFL096Cp [Eremothecium gossypii]	36577	1.9	53.8	1	6.90
15891262	AGR_L_2309p [Agrobacterium tumefaciens str. C58]	283038	0.2	53.8	1	5.22
22325705	hypothetical protein [Arabidopsis thaliana]	12857	5.4	53.8	1	9.95
34763159	hypothetical protein [Fusobacterium nucleatum subsp. vincentii ATCC 49256]	42809	1.6	53.8	1	5.20
29142237	membrane protein, suppressor for copper-sensitivity B precursor [Salmonella enterica subsp. enterica serovar Typhi Ty2]	68552	1.0	53.8	1	8.59
16759992	membrane protein, suppressor for copper-sensitivity B precursor [Salmonella enterica subsp. enterica serovar Typhi]	68580	1.0	53.8	1	8.59
15641108	oligopeptide ABC transporter, ATP-binding protein [Vibrio cholerae O1 biovar eltor str. N16961]	37061	1.8	53.8	1	8.95
17937390	peptide synthetase [Agrobacterium tumefaciens str. C58]	237031	0.3	53.8	1	5.18
16764472	suppression of copper sensitivity protein [Salmonella typhimurium LT2]	68623	1.0	53.8	1	8.75
16041069	glycogen branching enzyme [Aspergillus oryzae]	78701	1.9	53.8	3	5.62
29248824	GLP_22_45903_36058 [Giardia lamblia ATCC 50803]	370869	0.4	53.8	4	7.68
34540392	cell division protein FtsA [Porphyromonas gingivalis W83]	53875	3.1	53.7	3	5.55
2209034	FtsA [Porphyromonas gingivalis]	51193	3.3	53.7	3	5.54
15609566	ahpD [Mycobacterium tuberculosis H37Rv]	18769	9.0	53.7	3	6.51
24159058	Chain A, Crystal Structure Of Mycobacterium Tuberculosis Alkylperoxidase Ahpd H132q Mutant	18760	9.0	53.7	3	6.41
22219311	Chain A, Crystal Structure Of Mycobacterium Tuberculosis Alkylperoxidase Ahpd H137f Mutant	18779	9.0	53.7	3	6.41
23509685	chloroquine resistance marker protein [Plasmodium falciparum 3D7]	442796	0.4	53.6	4	8.64
46112098	COG1131: ABC-type multidrug transport system, ATPase component [Moorella thermoacetica ATCC 39073]	26087	4.3	53.5	2	6.21
15626070	vitellogenin [Larus argentatus]	201549	0.8	53.5	4	9.18
7441379	hypothetical protein C44B11.3 - Caenorhabditis elegans	50866	4.6	53.5	2	5.05
39580244	Hypothetical protein CBG20310 [Caenorhabditis briggsae]	50060	4.7	53.5	2	5.00
17554312	MEChanosensory abnormality MEC-12, TuBulin, Alpha, specific of 15 protofilament microtubules found in mechanosensory neurons (50.1 kD) (mec-12) [Caenorhabditis elegans]	50081	4.7	53.5	2	5.00

21064813	RH71862p [Drosophila melanogaster]	50034	4.7	53.5	2	5.10
39996908	cell division protein FtsH [Geobacter sulfurreducens PCA]	67516	1.5	53.4	2	5.91
46364347	COG0673: Predicted dehydrogenases and related proteins [Kineococcus radiotolerans SRS30216]	35510	2.4	53.4	1	5.00
29346817	activator of (R)-2-hydroxyglutaryl-CoA dehydratase [Bacteroides thetaiotaomicron VPI-5482]	156145	0.8	53.3	3	5.71
13358122	unique hypothetical [Ureaplasma urealyticum]	216587	0.8	53.1	4	8.70
27378001	phenylacetic acid degradation operon negative regulatory protein [Bradyrhizobium japonicum]	31669	2.4	53.1	1	6.62
7434407	S-receptor kinase (EC 2.7.1.-) - turnip	95772	1.3	53.1	3	8.22
1402512	S-receptor kinase SRK9 [Brassica rapa]	95513	1.3	53.1	3	8.11
23050378	Glycosyl transferases, related to UDP-glucuronosyltransferase [Methanosarcina barkeri]	42294	3.2	53.0	4	9.29
3334157	Peptidyl-prolyl cis-trans isomerase (PPIase) (Rotamase) (Cyclophilin) (Cyclosporin A-binding protein)	18273	8.7	53.0	2	8.36
29349208	two-component system sensor histidine kinase/response regulator, hybrid ('one-component system') [Bacteroides thetaiotaomicron VPI-5482]	151234	1.6	52.9	5	6.03
40746570	hypothetical protein AN0320.2 [Aspergillus nidulans FGSC A4]	165883	0.9	52.9	3	8.85
34912926	carboxypeptidase precursor-like protein [Oryza sativa (japonica cultivar-group)]	48047	1.5	52.8	1	6.62
34541249	lipoprotein signal peptidase, putative [Porphyromonas gingivalis W83]	25121	3.1	52.8	1	6.59
17551432	putative protein, with a transmembrane domain (XO70) [Caenorhabditis elegans]	23336	3.5	52.8	1	5.54
38503269	30S ribosomal protein S1	63801	3.0	52.7	4	6.52
32413959	hypothetical protein [Neurospora crassa]	20834	9.2	52.5	2	8.94
11465738	NP_053882.1 | ORF199 [Porphyra purpurea]	22268	7.5	52.5	2	4.68
44829140	TPA: ADP-ribosylation factor 1; ARF1 [Trypanosoma brucei]	20611	6.0	52.5	1	8.80
15595887	hypothetical protein [Pseudomonas aeruginosa PA01]	429760	0.3	52.4	2	4.72
15027208	hypothetical protein [Erwinia chrysanthemi]	112491	1.4	52.3	4	8.65
13021648	peptidyl-prolyl cis-trans isomerase B [Xenopus laevis]	11343	10.6	52.3	3	9.36
32484306	Ppib-prov protein [Xenopus laevis]	23845	5.1	52.3	3	9.30
39996920	N-acetylmuramoyl-L-alanine amidase, family 3 [Geobacter sulfurreducens PCA]	45687	1.4	52.2	1	9.00
28317324	AT04875p [Drosophila melanogaster]	141444	0.6	52.1	1	8.20
46318277	COG0827: Adenine-specific DNA methylase [Burkholderia cepacia R1808]	145858	1.0	52.0	3	5.12
37287387	neural alfa2 tubulin [Paracentrotus lividus]	50161	4.6	51.8	2	4.90
32400736	putative alpha-tubulin [Oikopleura dioica]	39662	5.9	51.8	2	5.00
39588311	Hypothetical protein CBG19875 [Caenorhabditis briggsae]	50009	4.8	51.8	2	5.03
23037696	ABC-type metal ion transport system, ATPase component [Oenococcus oeni MCW]	38957	4.5	51.8	2	9.07
42523406	cell division protein [Bdellovibrio bacteriovorus HD100]	70941	1.4	51.7	2	6.26
19924051	cytokine-induced neutrophil chemoattractant-2 [Rattus norvegicus]	10982	10.0	51.7	3	9.43
38090277	RIKEN cDNA 1110008I14 [Mus musculus]	196393	0.6	51.7	3	6.61
15594512	outer membrane protein (tpn50) [Borrelia burgdorferi]	45861	3.1	51.6	3	9.16
10179990	AAG13968.1 | cyclophilin [Magnaporthe grisea]	23538	4.2	51.6	1	9.86
10179991	AAG13969.1 | cyclophilin [Magnaporthe grisea]	17834	5.5	51.6	1	8.67
41615186	NEQ399 [Nanoarchaeum equitans Kin4-M]	32302	3.5	51.5	2	6.27

31615504	Chain A, Structure Of Deoxygenated Hemocyanin From Rapana Thomasiana	46707	2.7	51.5	1	5.02
31076723	Hemocyanin type 2 unit e (RtH2-e)	47933	2.7	51.5	1	5.17
38345704	OSJNBb0085C12.6 [Oryza sativa (japonica cultivar-group)]	128580	1.5	51.5	3	5.60
15922739	195aa long hypothetical protein [Sulfolobus tokodaii]	22818	5.1	51.4	3	8.67
39584414	Hypothetical protein CBG19736 [Caenorhabditis briggsae]	22088	4.4	51.4	1	8.85
16943775	putative cyclophilin [Pleurotus ostreatus]	17604	5.5	51.4	1	9.37
2497858	Malate dehydrogenase, mitochondrial precursor	38485	2.9	51.2	1	8.76
45768386	D230019K20Rik protein [Mus musculus]	163109	0.6	51.1	3	6.39
14779065	KIAA1093 protein [Homo sapiens]	182703	0.5	51.1	3	6.57
9629470	IVa2 protein [Duck adenovirus A]	45730	2.5	51.0	2	8.95
23482828	dynein beta chain, ciliary [Plasmodium yoelii yoelii]	604279	0.4	51.0	6	5.47
4138173	allergen [Malassezia sympodialis]	17202	5.6	49.7	1	8.79
547682	Heat shock cognate protein (Aginactin)	70456	2.2	49.0	2	5.40
13537429	small GTPase Rab1 [Entamoeba histolytica]	22882	7.3	49.0	2	5.95
45358795	conserved hypothetical protein [Methanococcus maripaludis S2]	34403	3.0	47.8	2	9.05
1708308	HEAT SHOCK PROTEIN 70 2	70030	2.2	45.7	2	4.95
124212	Type III intermediate filament	52787	3.3	44.3	3	5.06
40737034	putative integrase [Oryza sativa (japonica cultivar-group)]	171484	0.8	40.8	3	7.37
281428	genome polyprotein - soybean mosaic virus (strain G7)	349484	0.6	40.0	4	7.75
15807795	aldehyde dehydrogenase [Deinococcus radiodurans]	52254	2.2	38.4	2	4.86
41147635	protein phosphatase 1, regulatory (inhibitor) subunit 9A [Homo sapiens]	148111	0.8	37.5	3	5.30
38347275	OSJNBa0056L23.20 [Oryza sativa (japonica cultivar-group)]	72341	1.9	37.0	3	8.88
37534234	hypothetical protein [Oryza sativa (japonica cultivar-group)]	160432	0.8	36.5	3	7.53
46114676	hypothetical protein FG03180.1 [Gibberella zeae PH-1]	22746	5.4	35.9	2	4.94
17224953	tryparedoxin peroxidase [Trypanosoma cruzi]	22246	5.5	35.9	2	5.96
16798799	putative tape-measure protein [Bacteriophage A118]	186141	0.6	35.8	3	9.75
11131581	60 kDa chaperonin 1 (Protein Cpn60 1) (groEL1 protein)	57274	1.7	32.5	3	4.90

8.3 Appendix 3: Search result Conus textile in-gel

Accession Number	Protein Description	Mass (Da)	Sequence Coverage	Mascot Protein Score	Number of Peptides	pI
27817324	SI:zC101N13.3 (novel protein similar to human RAS oncogene family member RAB1B) [Danio rerio]	22274	38.8	340.8	11	5.56
15277503	ACTB protein [Homo sapiens]	40194	22.5	337.7	10	5.55
2829750	ACTIN	41297	21.8	337.7	10	5.46
17647135	CG10067-PA [Drosophila melanogaster]	41808	21.5	337.7	10	5.23
17137090	CG18290-PA [Drosophila melanogaster]	41775	21.5	337.7	10	5.30
17975545	CG5178-PA [Drosophila melanogaster]	41673	21.5	337.7	10	5.29
576368	Chain A, Beta-Actin-Profilin Complex	41801	21.5	337.7	10	5.21
14278147	Chain A, Crystal Structure Of Caenorhabditis Elegans Mg-Atp Actin Complexed With Human Gelsolin Seg	41638	21.6	337.7	10	5.30
31196067	ENSANGP00000016398 [Anopheles gambiae]	41986	21.4	337.7	10	5.38
31210881	ENSANGP00000015027 [Anopheles gambiae]	42369	21.1	337.7	10	5.47
31210009	ENSANGP00000009996 [Anopheles gambiae]	42696	21.0	337.7	10	5.48
31210879	ENSANGP00000015039 [Anopheles gambiae]	43637	20.5	337.7	10	5.47
31212569	ENSANGP00000022308 [Anopheles gambiae]	43653	20.5	337.7	10	5.47
31204457	ENSANGP00000019055 [Anopheles gambiae]	45438	19.8	337.7	10	5.46
809561	gamma-actin [Mus musculus]	40992	22.0	337.7	10	5.56
19528317	LD04994p [Drosophila melanogaster]	40232	22.5	337.7	10	5.56
45361511	hypothetical protein MGC75587 [Xenopus tropicalis]	41738	21.6	336.0	10	5.30
231498	ACTIN 2	41792	21.5	335.0	10	5.30
17530805	CG4027-PB [Drosophila melanogaster]	41795	21.5	335.0	10	5.30
3421457	cytoplasmic actin [Dreissena polymorpha]	41826	21.5	335.0	10	5.30
2624850	Chain A, Structure Of Bovine Beta-Actin-Profilin Complex With Actin Bound Atp Phosphates Solvent Ac	41664	21.6	334.3	10	5.15
6680602	keratin complex 1, acidic, gene 15; keratin 15 [Mus musculus]	49129	11.7	329.4	9	4.78
34784398	Krt1-15 protein [Mus musculus]	49463	11.6	329.4	9	4.79
8895859	actin [Cranchia scabra]	28519	35.2	312.7	10	5.14
13111394	larval keratin XLK [Xenopus laevis]	59904	8.6	302.3	7	5.82
6754488	keratin complex 2, basic, gene 6b [Mus musculus]	60285	9.6	292.3	7	8.51
3219772	ACTIN 51	37151	22.0	292.0	8	5.28
6678643	keratin complex 2, basic, gene 1 [Mus musculus]	65183	5.3	289.3	6	8.97
6754480	keratin complex 1, acidic, gene 13; keratin 13 [Mus musculus]	47724	10.8	284.8	9	4.79
23463261	vitamin K-dependent carboxylase [*Conus textile*]	93905	7.8	272.6	6	5.37
309215	EndoA" cytokeratin (5" end put.); putative	53210	6.3	252.8	5	5.42
13624315	keratin complex 2, basic, gene 8; cytokeratin 8; cytokeratin8; cytokeratin-8 [Mus musculus]	54220	6.2	252.8	5	5.40
19112997	ypt1-related protein 1 [Schizosaccharomyces pombe]	22802	32.0	242.9	8	5.31
32398899	small GTP binding protein rab1a, probable [Cryptosporidium parvum]	22790	24.1	228.5	7	7.63
1168322	ACTIN 71	41755	19.4	228.4	7	5.38
46275808	keratin complex 2, basic, gene 17; keratin complex 2, gene 17 [Mus musculus]	70934	7.2	220.5	6	8.23
38114739	KRT13 protein [Homo sapiens]	51029	9.3	217.2	7	4.95
49868	put. beta-actin (aa 27-375) [Mus musculus]	39161	17.8	216.2	7	5.78
38077190	expressed sequence AI507495 [Mus musculus]	56715	6.9	215.4	6	7.08

9910294	keratin complex 2, basic, gene 6g; keratin complex 2, gene 6g; caracul [Mus musculus]	57347	6.9	200.2	5	6.60
421942	GTP-binding protein, ras-related - common tobacco	22407	24.1	199.4	6	5.26
40786432	keratin complex 2, basic, gene 8; keratin 8 [Rattus norvegicus]	53985	7.0	194.0	5	5.83
349484	GTP-binding protein homologue	22210	19.5	193.8	5	5.55
3036959	CsCA1 [Ciona savignyi]	41698	14.1	193.8	7	5.29
27764565	TPA: keratin 5c [Homo sapiens]	59333	6.6	192.4	5	7.55
741022	keratin 15	49086	7.1	183.9	5	4.75
183817	beta-globin	18919	36.6	181.3	6	6.28
3660434	Chain B, Crystal Structure Of Deoxy-Human Hemoglobin Beta6 Glu->trp	15914	43.8	181.3	6	7.26
46014946	Chain B, Crystal Structure Of Human Hemoglobin E At 1.73 A Resolution	15856	43.8	181.3	6	7.98
3660145	Chain B, Crystal Structure Of S-Nitroso-Nitrosyl Human Hemoglobin A	15865	43.8	181.3	6	6.81
5822282	Chain B, Deoxygenated Structure Of A Distal Pocket Hemoglobin Mutant	15930	43.8	181.3	6	6.70
442854	Chain B, Hemoglobin (Deoxy) Mutant With Val 1 Replaced By Ala In The Beta Chains (Beta V1a)	15829	43.8	181.3	6	6.82
27065154	Chain B, Structure Of Mutant Human Carbonmonoxyhemoglobin C (Beta E6k) At 2.0 Angstrom Resolution I	15856	43.8	181.3	6	7.98
113226	ACTIN 15A	41800	10.9	181.1	6	5.29
71622	actin CyI - sea urchin (Strongylocentrotus purpuratus)	41804	10.9	181.1	6	5.29
7387473	CyI actin [Tripneustes gratilla]	41836	10.9	181.1	6	5.29
27525276	putative cytoskeletal actin [Ciona intestinalis]	41726	10.9	181.1	6	5.29
41018458	Substrate-specific endoprotease Tex31 precursor	33350	13.3	166.6	4	5.30
41393660	pupal-specific flight muscle actin [Aedes aegypti]	41572	8.8	165.3	5	5.29
114131	ADP-ribosylation factor	21751	14.1	163.1	5	8.48
29245482	GLP_334_11456_12031 [Giardia lamblia ATCC 50803]	21751	14.1	163.1	5	8.48
229149	hemoglobin beta	15866	32.9	160.6	5	5.23
3318722	Chain E, Leech-Derived Tryptase InhibitorTRYPSIN COMPLEX	23457	13.5	158.2	5	8.26
71626	actin 7 - fruit fly (Drosophila melanogaster)	41778	14.1	153.5	5	5.30
17975540	CG7478-PA [Drosophila melanogaster]	41760	14.1	153.5	5	5.30
21064361	RE12057p [Drosophila melanogaster]	37816	15.5	153.5	5	5.36
1703124	ACTIN 6	41497	14.2	151.8	5	5.31
85358	actin - sea urchin (Strongylocentrotus purpuratus)	41025	14.3	150.2	5	5.97
1703122	ACTIN 5	41157	14.3	150.2	5	5.56
31212573	ENSANGP00000022307 [Anopheles gambiae]	30160	19.4	150.2	5	5.13
728798	ACTIN 11	41762	14.1	147.5	5	5.24
17647133	CG12051-PA [Drosophila melanogaster]	41797	14.1	147.5	5	5.30
31212571	ENSANGP00000022306 [Anopheles gambiae]	41990	11.1	144.9	5	5.52
31241979	ENSANGP00000022175 [Anopheles gambiae]	42640	10.9	144.9	5	5.58
41393662	pupal-specific flight muscle actin [Aedes aegypti]	41556	11.2	144.9	5	5.29
5669640	ethylene-responsive small GTP-binding protein [Lycopersicon esculentum]	23755	13.9	144.3	4	8.37
1362065	GTP-binding protein GTP11 - garden pea	23687	14.0	144.3	4	7.66
1362067	GTP-binding protein GTP13 - garden pea	23791	14.0	144.3	4	7.66
1362066	GTP-binding protein GTP6 - garden pea	23924	13.9	144.3	4	7.63
479442	GTP-binding protein ypt2 - tomato	23896	13.9	144.3	4	7.64
33146687	putative ethylene-responsive small GTP-binding protein [Oryza sativa (japonica cultivar-group)]	23978	14.0	144.3	4	7.63

18447913	ras-related protein RAB8-1 [Nicotiana tabacum]	23959	13.9	144.3	4	7.66
7438394	small GTP-binding protein rab-1 - turnip	25005	13.5	144.3	4	6.17
31210883	ENSANGP00000015031 [Anopheles gambiae]	42932	10.9	141.6	5	5.32
18656777	ras-like protein [Pythium spinosum]	9731	31.0	136.1	5	6.92
20911031	keratin complex 2, basic, gene 5; tuftelin interacting protein 8; keratin 5 [Mus musculus]	61729	5.0	135.2	3	7.59
31559819	keratin 25C; type I inner root sheath specific keratin 25 irs3 [Homo sapiens]	49793	5.9	134.2	4	5.06
20520618	Keratin 12 [Oncorhynchus mykiss]	51904	4.8	134.0	4	5.15
20520616	Keratin 13 [Oncorhynchus mykiss]	51938	4.9	134.0	4	5.17
12214171	putative small GTP-binding protein (rab1b) [Homo sapiens]	18648	19.6	128.7	4	9.87
32527715	Ac2-048 [Rattus norvegicus]	33643	9.6	123.1	4	9.72
103720	GTP-binding protein o-rab1 - electric ray (Discopyge ommata)	22218	14.4	123.1	4	5.54
4585808	putative Rab1A protein [Plasmodium falciparum]	23814	11.1	121.4	3	7.60
37360618	mKIAA3012 protein [Mus musculus]	27317	11.5	121.4	4	7.68
7546413	Chain A, The Yeast Actin Val 159 Asn Mutant Complex With Human Gelsolin Segment 1	41678	6.4	120.4	4	5.44
14318479	Structural protein involved in cell polarization, endocytosis, and other cytoskeletal functions; Ac	41663	6.4	120.4	4	5.44
37928140	Chain A, Crystal Structure Of Oxy-Human Hemoglobin Bassett At 2.15 Angstrom	15073	28.4	119.2	4	9.07
3114508	Chain A, R State Human Hemoglobin [alpha V96w], Carbonmonoxy	15204	28.4	119.2	4	8.73
27574247	Chain A, Deoxy Hemoglobin (A,C:v1m,V62l; B,D:v1m,V67l)	15163	28.4	117.5	4	8.72
3891367	Chain A, Hemoglobin (Alpha V1m) Mutant	15149	28.4	117.5	4	8.72
996122	Chain A, Hemoglobin A (Deoxy) Mutant With Arg A 141 And Arg C 141 Deleted (Del R(A 141), R(C 141))	14961	28.6	117.5	4	8.07
493850	Chain A, Hemoglobin Thionville Alpha Chain Mutant With Val 1 Replaced By Glu And An Acetylated Met	15389	28.0	116.5	4	7.82
2982013	Chain A, Cyanomet Rhb1.1 (Recombinant Hemoglobin)	30305	14.1	112.4	4	8.91
9256890	Chain A, Deoxy Rhb1.2 (Recombinant Hemoglobin)	30362	14.1	112.4	4	8.91
122359	Hemoglobin alpha-I and alpha-II chains	14999	28.4	109.0	4	8.91
42523784	thiosulfate sulfurtransferase [Bdellovibrio bacteriovorus HD100]	31343	4.3	103.6	1	9.45
440306	enhancer protein	22113	8.5	102.5	2	8.16
7963723	natural killer cell enhancement factor [Oncorhynchus mykiss]	22001	8.5	102.5	2	6.51
27806081	peroxiredoxin 1 [Bos taurus]	22151	8.5	102.5	2	8.81
45185452	ABR220Wp [Eremothecium gossypii]	22737	13.7	100.9	3	5.04
16974365	AT4g17530/dl4800c [Arabidopsis thaliana]	22272	13.9	100.9	3	5.27
24648682	CG3320-PA [Drosophila melanogaster]	22748	13.7	100.9	3	5.31
28572143	CG3320-PB [Drosophila melanogaster]	16604	19.2	100.9	3	5.39
31208125	ENSANGP00000011746 [Anopheles gambiae]	23571	13.2	100.9	3	5.32
7438375	GTP-binding protein - Arabidopsis thaliana	24466	12.7	100.9	3	5.89
24306110	GTP-binding protein [Pichia angusta]	22561	13.8	100.9	3	5.48
303750	GTP-binding protein [Pisum sativum]	22484	13.9	100.9	3	5.28
303732	GTP-binding protein [Pisum sativum]	22449	13.8	100.9	3	5.11
81634	GTP-binding protein ara-5 - Arabidopsis thaliana (fragment)	21708	14.3	100.9	3	5.30
92339	GTP-binding protein rab1B - rat	22176	13.9	100.9	3	5.55
7438439	GTP-binding protein SYPT - soybean	23997	12.8	100.9	3	4.94
466171	GTP-binding protein ypt1	22462	13.8	100.9	3	5.47
82803	GTP-binding protein ypt1 - fission yeast (Schizosaccharomyces pombe)	23145	13.6	100.9	3	5.92

32418088	GTP-BINDING PROTEIN YPT1 [Neurospora crassa]	26693	11.5	100.9	3	6.82
2500073	GTP-binding protein YPTC1	22584	13.8	100.9	3	5.94
466172	GTP-binding protein YPTM2	22461	13.8	100.9	3	5.88
401686	GTP-binding protein yptV1	22518	13.8	100.9	3	5.93
541978	GTP-binding protein, ras-like (clone vfa-ypt1) - fava bean	22554	13.9	100.9	3	5.01
45360691	hypothetical protein MGC76044 [Xenopus tropicalis]	23083	13.9	100.9	3	8.21
38109427	hypothetical protein MG06962.4 [Magnaporthe grisea 70-15]	22448	13.9	100.9	3	5.27
34784624	MGC69101 protein [Xenopus laevis]	23110	13.9	100.9	3	8.52
34906164	putative GTP-binding protein [Oryza sativa (japonica cultivar-group)]	50107	6.2	100.9	3	9.83
21592670	putative Ras-like GTP-binding protein [Arabidopsis thaliana]	24070	13.0	100.9	3	7.63
38175435	putative ras-related protein [Oryza sativa (japonica cultivar-group)]	24026	13.0	100.9	3	7.68
34914060	putative RIC1_ORYSA RAS-RELATED PROTEIN RIC1 [Oryza sativa (japonica cultivar-group)]	22618	13.8	100.9	3	5.33
11558649	secretion related GTPase (SrgB) [Aspergillus niger]	22289	13.9	100.9	3	5.14
1053063	small GTP-binding protein	22542	13.8	100.9	3	5.89
1053067	small GTP-binding protein	22604	13.8	100.9	3	5.27
217841	small GTP-binding protein [Arabidopsis thaliana]	21609	14.4	100.9	3	5.30
7643790	small GTP-binding protein [Capsicum annuum]	22577	13.8	100.9	3	5.27
2808638	small GTP-binding protein [Daucus carota]	23940	13.0	100.9	3	7.63
1381678	small GTP-binding protein [Glycine max]	22410	13.9	100.9	3	5.12
1616614	small GTP-binding protein [Nicotiana plumbaginifolia]	22502	13.8	100.9	3	5.50
15077428	small GTP-binding protein Ypt1p [Candida albicans]	23005	13.5	100.9	3	5.27
11558500	small GTP-binding protein YPTI [Hypocrea jecorina]	22407	13.9	100.9	3	5.47
40739629	YPT1_NEUCR GTP-binding protein ypt1 [Aspergillus nidulans FGSC A4]	22204	14.0	100.9	3	5.53
7533034	YptA [Aspergillus awamori]	22321	13.9	100.9	3	5.14
7438417	GTP-binding protein rabB - silkworm	22350	13.9	99.2	3	5.44
39588485	Hypothetical protein CBG01077 [Caenorhabditis briggsae]	22561	13.7	99.2	3	5.52
17558550	RAB family member (22.5 kD) (rab-1) [Caenorhabditis elegans]	22531	13.7	99.2	3	5.52
2500076	Ras-like GTP-binding protein YPT1	22278	13.9	99.2	3	7.60
40738692	ACT_BOTCI Actin [Aspergillus nidulans FGSC A4]	41682	9.6	97.7	3	5.54
46125915	ACTG_CEPAC Actin, gamma [Gibberella zeae PH-1]	42180	9.5	97.7	3	5.54
202547	iodothyronine 5"-monodeiodinase	30136	4.9	97.0	4	5.09
6063529	70 kDa heat shock protein [Capra hircus]	69824	3.0	96.1	2	5.51
11177910	testis-specific heat shock protein-related gene hst70 [Rattus norvegicus]	69486	3.0	96.1	2	5.44
38103539	hypothetical protein MG03982.4 [Magnaporthe grisea 70-15]	39700	10.1	96.0	3	5.62
45382561	GTP-binding protein [Gallus gallus]	23507	11.8	94.5	2	5.98
106185	GTP-binding protein Rab2 - human	23549	11.8	94.5	2	6.08
28556900	GTP-binding protein rab-2 homologue [Ciona intestinalis]	20015	14.2	94.5	2	6.92
18652793	vitamin K-dependent gamma-glutamyl carboxylase [*Conus textile*]	93858	2.8	93.2	2	5.37
132270	Rubber elongation factor protein (REF) (Allergen Hev b 1)	14713	13.0	90.6	2	5.04
40787691	LOC395048 protein [Xenopus tropicalis]	56745	3.5	90.4	3	4.73
32822847	MGC64309 protein [Xenopus laevis]	56604	3.6	90.4	3	4.72
68288	carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)	28963	11.6	90.1	2	7.93
30466252	carbonic anhydrase II [Bos taurus]	29096	11.5	90.1	2	6.41
46016008	Chain A, Crystal Structure Analysis Of Bovine Carbonic Anhydrase Ii	28965	11.6	90.1	2	6.40
46016010	Chain C, Crystal Structure Analysis Of The Site Specific Mutant	29199	11.5	90.1	2	6.63

	(Q253c) Of Bovine Carbonic Anhydras					
4557581	fatty acid binding protein 5 (psoriasis-associated) [Homo sapiens]	15155	13.3	89.4	2	6.60
28436918	P4hb protein [Xenopus laevis]	57980	3.5	88.7	3	4.81
2914482	Chain A, Complex Of The Second Kunitz Domain Of Tissue Factor Pathway Inhibitor With Porcine Trypsi	23460	13.5	88.6	5	8.26
1942351	Chain A, Crystal Structure Of The First Active Autolysate Form Of The Porcine Alpha Trypsin	13284	24.0	88.6	5	7.83
494360	Chain A, Trypsin (E.C.3.4.21.4) Complexed With Inhibitor From Bitter Gourd	23458	13.5	88.6	5	8.60
999627	Chain B, Porcine E-Trypsin (E.C.3.4.21.4)	8814	36.6	88.6	5	6.67
3183057	ARGININE KINASE (AK)	39077	9.5	87.7	3	7.12
24585671	CG17949-PA [Drosophila melanogaster]	13688	18.7	85.1	3	10.34
31195811	ENSANGP00000000106 [Anopheles gambiae]	10626	24.2	85.1	3	9.79
31194099	ENSANGP00000012781 [Anopheles gambiae]	11308	23.0	85.1	3	10.22
31239833	ENSANGP00000014150 [Anopheles gambiae]	13751	18.5	85.1	3	10.34
31196269	ENSANGP00000012046 [Anopheles gambiae]	14426	17.7	85.1	3	10.27
31194615	ENSANGP00000000003 [Anopheles gambiae]	15691	16.2	85.1	3	10.46
31239831	ENSANGP00000014080 [Anopheles gambiae]	17000	15.0	85.1	3	10.48
17561988	histone (his-4) [Caenorhabditis elegans]	15634	16.3	85.1	3	10.40
38564173	histone 2B [Drosophila sechellia]	13212	19.3	85.1	3	10.32
39581593	Hypothetical protein CBG01507 [Caenorhabditis briggsae]	13492	18.9	85.1	3	10.40
39583631	Hypothetical protein CBG10818 [Caenorhabditis briggsae]	13535	18.7	85.1	3	10.34
17562020	predicted CDS, histone (his-39) [Caenorhabditis elegans]	12702	20.5	85.1	3	9.99
6686553	putative H2B histone [Asellus aquaticus]	13764	18.7	85.1	3	10.39
37589438	MGC69157 protein [Xenopus laevis]	25098	7.5	85.0	2	6.70
3023913	Putative RAS-related protein COS41.2	24602	7.8	85.0	2	6.84
20306864	Keratin 23, isoform a [Homo sapiens]	48147	4.3	84.2	3	6.09
22001960	40S ribosomal protein S15a	14735	13.1	83.9	2	9.89
24652557	CG12324-PA [Drosophila melanogaster]	14732	13.1	83.9	2	9.84
17975567	CG2033-PD [Drosophila melanogaster]	14762	13.1	83.9	2	9.84
36142	ribosomal protein homologous to yeast S24 [Homo sapiens]	14707	13.2	83.9	2	10.23
18000277	ribosomal protein S15 isoform [Lapemis hardwickii]	14846	13.1	83.9	2	10.07
45504881	ribosomal protein S15a [Danio rerio]	14830	13.1	83.9	2	10.14
12804561	Ribosomal protein S15a [Homo sapiens]	14770	13.1	83.9	2	10.14
1082767	ribosomal protein S15a, cytosolic [validated] - human	14844	13.1	83.9	2	10.14
14165469	ribosomal protein S15a; 40S ribosomal protein S15a [Homo sapiens]	14830	13.1	83.9	2	10.14
30109302	Rps15a protein [Mus musculus]	15610	12.3	83.9	2	10.01
39595176	Hypothetical protein CBG03777 [Caenorhabditis briggsae]	103310	2.5	83.4	3	9.23
39595059	Hypothetical protein CBG17727 [Caenorhabditis briggsae]	20549	8.9	82.4	2	6.43
17136990	CG5409-PA [Drosophila melanogaster]	42003	6.4	81.8	3	5.63
12084563	Chain A, Crystal Structure Of Sec4-Gdp	19071	12.9	80.1	2	5.11
12084567	Chain A, Crystal Structure Of Sec4-Guanosine-5"-(Beta,Gamma)-Imidotriphosphate	19131	12.9	80.1	2	5.11
23613148	GTPase, putative [Plasmodium falciparum 3D7]	22872	11.0	80.1	2	6.18
39645075	MGC68722 protein [Xenopus laevis]	16554	15.4	80.1	2	8.81
23479748	putative GTPase [Plasmodium yoelii yoelii]	25890	9.8	80.1	2	8.46
14318517	Secretory vesicle associated Rab GTPase that binds to Sec15p and is essential for exocytosis; Sec4p [Saccharomyces cerevisiae]	23491	10.2	80.1	2	6.62
9965598	peroxidoxin-2 [Litomosoides sigmodontis]	21320	7.8	79.2	2	6.41

12751382	thioredoxin peroxidase [Brugia malayi]	18005	9.2	79.2	2	6.28
19698783	thiredoxin peroxidase [Acanthocheilonema viteae]	27703	6.1	79.2	2	5.10
2959746	putative SEC4 protein [Candida glabrata]	23666	10.2	78.4	2	6.62
3212313	Chain A, Artificial Mutant (Alpha Y42h) Of Deoxy Hemoglobin	15091	15.6	77.2	2	8.76
3212437	Chain A, R-State Human Carbonmonoxyhemoglobin Alpha-A53s	15133	15.6	77.2	2	8.73
28277305	Vim4 protein [Xenopus laevis]	53507	3.5	76.8	2	5.12
32450567	Vim4 protein [Xenopus laevis]	53572	3.5	76.8	2	5.12
138532	Vimentin 4	53464	3.5	76.8	2	5.08
138531	Vimentins 1 and 2	52812	3.5	76.8	2	5.16
21357783	CG7768-PB [Drosophila melanogaster]	17760	7.3	75.8	2	8.32
45549139	CG9916-PA [Drosophila melanogaster]	24650	5.3	75.8	2	9.34
32450581	LOC398630 protein [Xenopus laevis]	20896	6.2	75.8	2	9.12
229953	Chain G, D-Glyceraldehyde-3-Phosphate Dehydrogenase (E.C.1.2.1.12)	35804	4.2	74.6	1	6.29
9967445	glyceraldehyde-3-phosphate dehydrogenase [Actinomucor elegans]	8694	17.5	74.6	1	4.81
39589064	Hypothetical protein CBG00820 [Caenorhabditis briggsae]	36229	4.1	74.6	1	8.24
39581573	Hypothetical protein CBG01479 [Caenorhabditis briggsae]	36376	4.1	74.6	1	7.68
38603063	putative cytosolic glyceraldehyde-3-phosphate dehydrogenase [Cryptomeria japonica]	15563	9.6	74.6	1	7.90
14423267	small GTP-binding protein Rab8 [Entamoeba histolytica]	16478	19.3	74.4	3	5.58
14423577	small GTP-binding protein Rab8 [Entamoeba histolytica]	22258	14.1	74.4	3	5.05
17647301	CG4886-PA [Drosophila melanogaster]	33255	4.0	74.2	2	5.72
7770087	arginine kinase [Sepioteuthis lessoniana]	39059	5.2	74.1	2	6.28
11465738	ORF199 [Porphyra purpurea]	22268	7.5	73.0	2	4.68
16945685	disulfide isomerase [Ostertagia ostertagi]	54972	2.8	72.9	3	4.93
9506669	keratin 24 [Homo sapiens]	55111	3.6	72.3	2	4.86
17565064	histone 3.3 (15.3 kD) (5N140) [Caenorhabditis elegans]	15340	10.3	71.0	2	11.40
15238430	histone H3 [Arabidopsis thaliana]	15582	10.1	71.0	2	11.26
17137088	CG3269-PA [Drosophila melanogaster]	23546	5.2	70.7	1	5.98
6517192	Drab2 [Drosophila melanogaster]	23523	5.2	70.7	1	5.85
7438437	GTP-binding protein rab2 - rice	24903	4.9	70.7	1	8.74
7438383	GTP-binding protein rab2 - soybean	23155	5.2	70.7	1	7.70
7438379	GTP-binding protein RAB2A - Arabidopsis thaliana	22887	5.3	70.7	1	6.97
2500074	GTP-binding protein YPTC4	23583	5.2	70.7	1	7.01
7438438	GTP-binding protein yptm3 - maize	23027	5.2	70.7	1	6.96
549811	GTP-binding protein yptV4 (RAB2 homolog)	23682	5.2	70.7	1	7.01
39586261	Hypothetical protein CBG12011 [Caenorhabditis briggsae]	23635	5.1	70.7	1	6.22
38344743	OSJNBa0089K21.1 [Oryza sativa (japonica cultivar-group)]	23070	5.2	70.7	1	6.96
5738168	putative intermediate compartment protein [Tetrahymena thermophila]	23873	5.2	70.7	1	6.84
13397937	putative Rab2 GTPase [Plasmodium falciparum 3D7]	24394	5.2	70.7	1	6.33
16755592	small GTPase Rab2 [Nicotiana tabacum]	23155	5.2	70.7	1	6.90
6624302	small GTP-binding protein [Carica papaya]	23294	5.2	70.7	1	8.43
33357681	putative mitochondrial cyclophilin 1 [Botryotinia fuckeliana]	24210	8.1	70.6	2	9.14
39595509	Hypothetical protein CBG04174 [Caenorhabditis briggsae]	40331	6.9	70.2	3	8.58
13879585	1110014C03Rik protein [Mus musculus]	11327	9.6	70.2	1	7.82
17561130	emp24/gp25L/p24 family (23.7 kD) (5L804) [Caenorhabditis elegans]	23665	4.4	70.2	1	7.06
39592002	Hypothetical protein CBG23171 [Caenorhabditis briggsae]	23618	4.4	70.2	1	7.02
16758214	integral membrane protein Tmp21-I (p23) [Rattus norvegicus]	24842	4.1	70.2	1	6.02

32484216	MGC64264 protein [Xenopus laevis]	23761	4.4	70.2	1	5.81
32484308	MGC64308 protein [Xenopus laevis]	23976	4.3	70.2	1	6.92
6689664	p24 delta1 putative cargo receptor [Xenopus laevis]	23930	4.3	70.2	1	6.92
6323630	Protein that forms a heterotrimeric comnplex with Erp1, Erp2p, and Emp24, member of the p24 family involved in endoplasmic reticulum to Golgi transport; Erv25p [Saccharomyces cerevisiae]	24092	4.3	70.2	1	5.36
3087752	S31 protein [Cyprinus carpio]	14867	7.0	70.2	1	4.91
3915137	Transmembrane protein Tmp21 precursor (21 kDa Transmembrane trafficking protein)	23262	4.4	70.2	1	6.03
1063411	unnamed protein product [Xenopus laevis]	23771	4.4	70.2	1	5.69
37748485	Wu:fe06g04 protein [Danio rerio]	24426	4.3	70.2	1	6.37
23127071	ATPase involved in DNA repair [Nostoc punctiforme]	59730	2.3	69.5	2	5.06
202549	iodothyronine 5" monodeiodinase	54033	3.1	69.4	2	4.87
27806501	procollagen-proline, 2-oxoglutarate 4-dioxygenase [proline 4-hydroxylase], beta polypeptide [protei	57230	2.9	69.4	2	4.80
6981324	prolyl 4-hydroxylase, beta polypeptide; Protein disulfide isomerase (Prolyl 4-hydroxylase, beta po	56829	3.0	69.4	2	4.82
339647	thyroid hormone binding protein precursor	57069	3.0	69.4	2	4.82
24641150	CG32671-PA [Drosophila melanogaster]	22758	5.6	69.3	1	6.21
24641047	CG32678-PA [Drosophila melanogaster]	22830	5.6	69.3	1	6.21
20129057	CG9575-PA [Drosophila melanogaster]	22848	5.5	69.3	1	8.72
24641058	CG9807-PA [Drosophila melanogaster]	22934	5.6	69.3	1	5.76
4895063	GTP-binding protein [Trypanosoma cruzi]	23828	5.0	69.3	1	6.15
549813	GTP-binding protein yptV5	23038	5.4	69.3	1	5.73
39584751	Hypothetical protein CBG13205 [Caenorhabditis briggsae]	23329	5.3	69.3	1	8.54
6983543	LmRab7 GTP-binding protein [Leishmania major]	24134	4.9	69.3	1	5.27
33284843	SI:dZ146N9.2 (novel protein similar to human member of RAS oncogene family (RAB35)) [Danio rerio]	8552	14.5	69.3	1	8.09
227879	cyclophilin	18995	8.4	69.2	2	8.65
32449722	ACTN1 protein [Xenopus laevis]	102992	1.1	69.0	1	5.22
38018016	brain-specific alpha actinin 1 isoform [Rattus norvegicus]	105405	1.1	69.0	1	5.23
17511011	arginine kinase family member (40.4 kD) (1K924) [Caenorhabditis elegans]	40357	6.9	68.5	3	8.57
14602206	predicted transcription regulator [Aeropyrum pernix]	16148	10.6	68.3	2	9.86
24641064	CG32673-PA [Drosophila melanogaster]	22871	5.6	67.6	1	6.30
1666188	ORF8(2) [Rhodococcus erythropolis]	51619	2.5	67.6	2	8.83
46249431	protein disulfide isomerase [Ancylostoma caninum]	54749	3.2	67.6	2	4.88
38109352	hypothetical protein MG06894.4 [Magnaporthe grisea 70-15]	144266	0.9	67.5	2	5.41
63389	fibroblast alpha actinin [Gallus gallus]	98415	1.2	67.4	1	5.32
3551774	alpha-glycerophosphate oxidase [Streptococcus pneumoniae]	66751	2.1	67.2	2	4.98
15901992	hypothetical protein [Streptococcus pneumoniae TIGR4]	66755	2.1	67.2	2	4.98
29437296	Actg2 protein [Mus musculus]	22441	9.9	67.2	3	5.32
15229955	actin 11 (ACT11) [Arabidopsis thaliana]	41648	5.3	67.2	3	5.23
3219763	ACTIN 53	37346	6.0	67.2	3	5.39
1669389	actin 8	41862	5.3	67.2	3	5.37
6636384	alpha-cardiac actin [Danio rerio]	41969	5.3	67.2	3	5.29
294852	alpha-muscle actin	41511	5.4	67.2	3	5.47
15983789	AT3g18780/MVE11_16 [Arabidopsis thaliana]	38597	5.8	67.2	3	6.03
309090	A-X actin	41667	5.3	67.2	3	5.21
22725710	beta actin-like protein [Chelydra serpentina serpentina]	19351	11.7	67.2	3	5.49

7766848	Chain A, Complex Between Rabbit Muscle Alpha-Actin: Human Gelsolin Domain 1	41998	5.3	67.2	3	5.17
229690	Chain A, Deoxyribonuclease I Complex With Actin	41523	5.4	67.2	3	5.09
20664362	Chain A, Polylysine Induces An Antiparallel Actin Dimer That Nucleates Filament Assembly: Crystal S	41273	5.4	67.2	3	5.46
39654752	Chain A, Structure Of Rabbit Actin In Complex With Kabiramide C	41764	5.3	67.2	3	5.17
15825662	Chain A, Uncomplexed Actin	41875	5.3	67.2	3	5.10
15321578	fast muscle actin [Scyliorhinus retifer]	42002	5.3	67.2	3	5.16
39579170	Hypothetical protein CBG24445 [Caenorhabditis briggsae]	37606	5.9	67.2	3	5.71
39591126	Hypothetical protein CBG02161 [Caenorhabditis briggsae]	41059	5.4	67.2	3	5.09
39591127	Hypothetical protein CBG02162 [Caenorhabditis briggsae]	41087	5.4	67.2	3	5.09
39579169	Hypothetical protein CBG24444 [Caenorhabditis briggsae]	41597	5.3	67.2	3	5.20
45361557	hypothetical protein MGC75582 [Xenopus tropicalis]	42006	5.3	67.2	3	5.23
45360805	hypothetical protein MGC75697 [Xenopus tropicalis]	41957	5.3	67.2	3	5.22
28436790	MGC53823 protein [Xenopus laevis]	41992	5.3	67.2	3	5.22
2967678	smooth muscle gamma actin [Gallus gallus]	41876	5.3	67.2	3	5.31
4902905	unnamed protein product [Xenopus laevis]	42024	5.3	67.2	3	5.23
64509	unnamed protein product [Xenopus laevis]	41940	5.3	67.2	3	5.22
21357425	keratin complex 2, basic, gene 20 [Mus musculus]	57076	3.5	67.1	2	5.99
24158633	Chain A, Cyclophilin_5 From C. Elegans	19848	4.4	66.6	1	9.22
39584414	Hypothetical protein CBG19736 [Caenorhabditis briggsae]	22088	3.9	66.6	1	8.85
32484306	Ppib-prov protein [Xenopus laevis]	23845	3.7	66.6	1	9.30
212995	keratin	49724	3.3	65.5	2	5.27
29246691	GLP_9_1234_4455 [Giardia lamblia ATCC 50803]	120953	1.2	64.7	2	5.73
14245706	kinesin-like protein 6 [Giardia intestinalis]	109871	1.3	64.7	2	5.62
39581425	Hypothetical protein CBG22522 [Caenorhabditis briggsae]	135073	0.9	64.1	2	8.86
45185754	ACR068Wp [Eremothecium gossypii]	209157	0.8	64.0	2	6.24
15897817	Glutamyl-tRNA -Gln amidotransferase, subunit B (gatB-2) [Sulfolobus solfataricus]	71490	2.2	64.0	2	5.91
34763462	hypothetical protein [Fusobacterium nucleatum subsp. vincentii ATCC 49256]	96812	1.4	63.6	2	9.00
37362272	2-peptidylprolyl isomerase A [Danio rerio]	17390	9.1	63.5	2	8.87
45382027	S-cyclophilin [Gallus gallus]	22399	16.9	63.2	3	9.40
15606629	hypothetical protein [Aquifex aeolicus]	28500	4.8	62.8	2	10.16
20808349	Transcriptional regulator of sugar metabolism [Thermoanaerobacter tengcongensis]	27854	5.2	62.7	2	8.67
23110237	COG1012: NAD-dependent aldehyde dehydrogenases [Novosphingobium aromaticivorans]	53766	2.2	62.5	1	5.32
3334157	Peptidyl-prolyl cis-trans isomerase (PPIase) (Rotamase) (Cyclophilin) (Cyclosporin A-binding protein)	18273	8.7	62.1	2	8.36
41387680	phosphoenolpyruvate carboxylase [Chlamydomonas reinhardtii]	131160	1.4	62.1	2	6.10
15078743	030L [Invertebrate iridescent virus 6]	60163	1.3	62.0	1	8.33
17543744	6-phosphogluconolactonase (30.4 kD) (4P533) [Caenorhabditis elegans]	30395	2.6	62.0	1	5.16
23028343	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases [Microbulbifer degradans 2-40]	65729	1.1	62.0	1	5.07
15835008	ATP-dependent Clp protease, subunit B [Chlamydia muridarum]	96544	0.8	62.0	1	5.43
23017454	COG1201: Lhr-like helicases [Thermobifida fusca]	80264	1.0	62.0	1	6.39
20808304	hypothetical protein [Thermoanaerobacter tengcongensis]	13082	5.8	62.0	1	4.59
37523779	probable ABC transporter ATP-binding protein [Gloeobacter	34946	2.2	62.0	1	8.89

	violaceus PCC 7421]					
25309645	Y57G11C.3 [imported] - Caenorhabditis elegans	29057	2.7	62.0	1	5.26
27806469	peptidylprolyl isomerase B [Bos taurus]	22687	16.8	61.5	3	9.23
18025342	beta-D-xylosidase [Hordeum vulgare]	83476	2.6	61.4	2	6.36
42781880	acetyltransferase, GNAT family [Bacillus cereus ATCC 10987]	16641	6.9	61.2	2	5.84
14518283	C-type lectin CTL-2 precursor [Necator americanus]	19979	3.9	60.7	1	4.73
13647113	arginine kinase [Cellana grata]	38497	5.2	60.7	2	5.91
32562852	abnormal CHEmotaxis CHE-3, altered AVeRmectin sensitivity AVR-1, OSMotic avoidance abnormal OSM-2, abnormal CAFfeine-resistance CAF-2, dynein heavy chain (che-3) [Caenorhabditis elegans]	473519	0.2	60.3	1	5.81
37520551	hypothetical protein glr0982 [Gloeobacter violaceus PCC 7421]	126916	0.6	60.3	1	5.42
7499348	hypothetical protein F18C12.1 - Caenorhabditis elegans	469369	0.2	60.3	1	5.79
38344927	OSJNBa0018M05.18 [Oryza sativa (japonica cultivar-group)]	521760	0.2	60.3	1	5.70
17512384	BC031593 protein [Mus musculus]	59144	4.4	60.2	2	7.10
22164776	cDNA sequence BC031593 [Mus musculus]	57517	4.5	60.2	2	7.55
23613877	hypothetical protein [Plasmodium falciparum 3D7]	49524	4.6	60.2	2	4.75
6319146	H2A protein [Oryza sativa]	14485	6.5	59.9	1	10.18
15232536	histone H2A.F/Z [Arabidopsis thaliana]	14532	6.6	59.9	1	10.32
37534390	putative histone H2A [Oryza sativa (japonica cultivar-group)]	14603	6.5	59.9	1	10.39
23053281	COG0117: Pyrimidine deaminase [Geobacter metallireducens]	39502	2.2	59.8	1	9.16
439534	guanine nucleotide regulatory protein	14418	16.0	59.8	2	8.86
1834381	first subunit of lichenysin synthetase [Bacillus licheniformis]	402113	0.3	59.5	2	5.17
3080742	lichenysin synthetase A; LicA [Bacillus licheniformis]	402324	0.3	59.5	2	5.20
39582454	Hypothetical protein CBG11853 [Caenorhabditis briggsae]	515694	0.2	59.3	1	5.98
46132850	COG1249: Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Ralstonia eutropha JMP134]	49120	1.7	59.3	1	6.03
13182948	Msh4p-related protein [Candida glabrata]	101107	1.7	59.0	2	8.73
4929993	Chain A, Module-Substituted Chimera Hemoglobin Beta-Alpha (F133v)	15780	11.0	59.0	2	7.21
1942686	Chain B, (Alpha-Oxy, Beta-(C112g)deoxy) T-State Human Hemoglobin	15811	11.0	59.0	2	6.81
27574230	Chain B, Cyanomet Hemoglobin (A-Gly-C:v1m,L29f,H58q; B,D:v1m,L106w)	15962	11.0	59.0	2	6.75
2982014	Chain B, Cyanomet Rhb1.1 (Recombinant Hemoglobin)	15903	11.0	59.0	2	7.13
1942682	Chain B, Deoxy (Beta-(C93a,C112g)) Human Hemoglobin	15779	11.0	59.0	2	6.82
27574252	Chain B, Deoxy Hemoglobin (A-Gly-C:v1m; B,D:v1m,C93a,N108k)	15871	11.0	59.0	2	7.14
2981643	Chain B, Hemoglobin (Val Beta1 Met, Trp Beta37 Tyr) Mutant	15866	11.0	59.0	2	6.75
183855	hbbm fused globin protein (beta chain sequence to base 268)	10922	15.8	59.0	2	6.17
229172	hemoglobin delta	15897	11.0	59.0	2	7.97
13432059	mutant beta globin [Homo sapiens]	12226	14.4	59.0	2	5.65
30268354	hypothetical protein [Homo sapiens]	185707	1.5	59.0	2	5.87
24649452	CG5338-PB [Drosophila melanogaster]	17351	9.7	58.9	2	10.23
41054687	coated vesicle membrane protein [Danio rerio]	22915	4.5	58.8	1	5.08
1352660	Cop-coated vesicle membrane protein p24 precursor	22175	4.6	58.8	1	5.07
28277266	Rnp24-prov protein [Xenopus laevis]	22826	4.5	58.8	1	4.99
32399098	hypothetical predicted protein, unknown function	255333	0.8	58.7	2	5.44

	[Cryptosporidium parvum]					
15616868	GTP-binding protein Era [Buchnera aphidicola str. APS (Acyrthosiphon pisum)]	32881	4.6	58.5	2	9.77
40744755	hypothetical protein AN2226.2 [Aspergillus nidulans FGSC A4]	269269	0.3	58.4	1	8.91
15607114	transcriptional regulator (TetR/AcrR family) [Aquifex aeolicus]	22245	3.6	58.4	1	8.51
28573140	CG33104-PA [Drosophila melanogaster]	25143	5.1	58.3	2	6.55
28573133	CG33105-PA [Drosophila melanogaster]	28158	4.5	58.3	2	6.25
15920300	963aa long hypothetical protein [Sulfolobus tokodaii]	112099	1.2	58.2	2	8.19
39596886	Hypothetical protein CBG02408 [Caenorhabditis briggsae]	79728	1.8	58.1	2	5.31
2459908	anon2A5 [Drosophila melanogaster]	31665	3.3	57.9	2	6.13
45550758	CG10340-PA [Drosophila melanogaster]	32014	3.2	57.9	2	6.11
38233306	Translocase protein [Corynebacterium diphtheriae]	96033	1.9	57.9	2	4.91
205055	keratin K5	9390	15.2	57.8	1	5.34
15899583	3-hydroxyacyl-CoA dehydrogenase (hdb-2) [Sulfolobus solfataricus]	41452	3.0	57.8	2	6.81
6730537	A Chain A, Crystal Structure Of Vip2 With Nad	45462	3.0	57.7	2	6.26
41407942	Lpd [Mycobacterium avium subsp. paratuberculosis str. k10]	48706	1.8	57.6	1	6.33
7435435	phospho-beta-galactosidase I - Lactobacillus gasseri	55131	4.1	57.3	3	5.32
38105643	hypothetical protein MG03637.4 [Magnaporthe grisea 70-15]	176753	0.8	57.2	2	8.69
39579623	Hypothetical protein CBG22322 [Caenorhabditis briggsae]	41252	5.6	57.2	3	5.88
21223342	putative regulatory protein [Streptomyces coelicolor A3(2)]	124243	0.8	57.0	2	6.91
13310482	axonemal dynein heavy chain 8 Dnahc8 [Mus musculus]	397608	0.3	57.0	2	5.80
31088005	DNA polymerase B1 [Acidianus brierleyi]	100215	2.1	56.9	2	8.60
2982220	Rho-associated kinase alpha [Xenopus laevis]	159185	1.5	56.8	3	5.93
41386727	protein kinase C substrate 80K-H [Bos taurus]	60113	1.9	56.7	1	4.36
22970133	hypothetical protein [Chloroflexus aurantiacus]	78923	0.9	56.7	1	4.78
18640216	SPV130 DNA ligase-like ptorein [Swinepox virus]	64677	3.0	56.5	2	8.92
10799285	keratin 17 [Notophthalmus viridescens]	24351	6.6	56.4	2	5.14
1752728	alpha-aminoadipyl-cysteinyl-valine synthetase [Lysobacter lactamgenus]	411361	0.4	56.3	2	5.47
32363171	Histone H2B	12523	19.5	56.2	2	10.53
11466760	ORF464 [Marchantia polymorpha]	57130	4.7	56.1	2	10.32
15615079	fibronectin/fibrinogen-binding protein [Bacillus halodurans]	65143	4.2	56.0	2	8.97
23482768	chloroquine resistance marker protein [Plasmodium yoelii yoelii]	425780	0.3	55.8	2	8.95
7486705	hypothetical protein F8M12.17 - Arabidopsis thaliana	182928	1.3	55.8	2	8.84
37524286	hypothetical protein [Photorhabdus luminescens subsp. laumondii TTO1]	28560	3.8	55.7	2	7.82
18418335	armadillo/beta-catenin repeat family protein [Arabidopsis thaliana]	49758	1.5	55.6	1	4.82
10185816	auxin-induced protein TGSAUR12 [tulipa gesneriana]	11340	7.1	55.6	1	9.26
7485612	hypothetical protein F17I5.130 - Arabidopsis thaliana	73367	1.0	55.6	1	6.22
29831315	putative DJ-1/PfpI-family transcriptional regulator [Streptomyces avermitilis MA-4680]	23861	3.0	55.6	1	4.77
34557268	TRANSCRIPTIONAL REGULATOR (LYSR FAMILY) [Wolinella succinogenes]	34548	2.3	55.6	1	8.91
2764623	95 kD basal apparatus-protein [Spermatozopsis similis]	90194	1.1	55.6	1	7.44
23477111	histone H4 [Colletotrichum sp.]	6814	16.7	55.5	1	10.67
19851831	fatty acid synthase beta subunit [Aspergillus parasiticus]	210297	0.7	55.4	3	6.51
18146786	CYP1 [Vigna radiata]	18195	8.7	55.1	2	8.70
6323562	Mitochondrial peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; involved in protein refolding after import into	19906	8.2	55.1	2	8.81

	mitochondria; Cpr3p [Saccharomyces cerevisiae]					
18376589	putative cyclosporin A-binding protein [Picea abies]	18255	8.7	55.1	2	8.39
15927529	hypothetical protein [Staphylococcus aureus subsp. aureus N315]	143071	1.1	55.1	2	5.07
23039735	COG0675: Transposase and inactivated derivatives [Trichodesmium erythraeum IMS101]	41298	3.0	55.0	3	9.94
17551432	putative protein, with a transmembrane domain (XO70) [Caenorhabditis elegans]	23336	3.5	54.9	1	5.54
7682378	DNA polymerase [Leishmania enriettii]	34661	4.9	54.8	4	8.78
23509787	hypothetical protein [Plasmodium falciparum 3D7]	113162	0.9	54.8	2	5.51
46102596	hypothetical protein FG00002.1 [Gibberella zeae PH-1]	75376	2.8	54.6	2	5.18
729278	PPIA_STRCH Peptidyl-prolyl cis-trans isomerase A (PPIase A) (Rotamase A) (Cyclophilin ScCypA) (Cyclophilin homolog)	17705	9.1	54.5	2	5.32
46096722	hypothetical protein UM01171.1 [Ustilago maydis 521]	61573	3.1	54.4	2	7.96
15900638	transcriptional regulator, TetR family [Streptococcus pneumoniae TIGR4]	21714	5.4	54.4	1	9.35
21358147	CG2358-PA [Drosophila melanogaster]	20967	4.3	54.3	1	8.80
28829085	hypothetical protein [Dictyostelium discoideum]	20132	4.5	54.3	1	6.51
23028043	COG1115: Na+/alanine symporter [Microbulbifer degradans 2-40]	47221	1.5	54.2	1	9.33
41152179	hypothetical protein MGC73211 [Danio rerio]	15932	8.2	54.0	2	10.23
10121615	ribosomal protein S19 [Gillichthys mirabilis]	16176	8.2	54.0	2	10.18
16924231	ribosomal protein S19 [Homo sapiens]	17271	7.6	54.0	2	10.52
12963511	ribosomal protein S19 [Mus musculus]	16076	8.3	54.0	2	10.41
37779108	ribosomal protein S19 [Pagrus major]	16035	8.2	54.0	2	10.22
34855342	ribosomal protein S19 [Rattus norvegicus]	54950	2.4	54.0	2	10.45
4506695	ribosomal protein S19; 40S ribosomal protein S19 [Homo sapiens]	16051	8.3	54.0	2	10.31
34194029	Rps19-prov protein [Xenopus laevis]	16040	8.2	54.0	2	10.25
46370702	cytochrome P450 [Ammi majus]	59689	1.3	53.7	1	8.28
15923083	hypothetical protein [Staphylococcus aureus subsp. aureus Mu50]	121704	1.0	53.5	2	5.81
32029187	COG5295: Autotransporter adhesin [Haemophilus somnus 2336]	225562	1.0	53.5	2	6.52
32401451	hypothetical protein FLJ90492 [Homo sapiens]	103769	1.6	53.2	2	9.03
6225121	60 kDa chaperonin (Protein Cpn60) (groEL protein)	56623	4.8	53.2	1	4.85
33087653	60 kDa heat shock protein [Yersinia rohdei]	10449	27.1	53.2	1	4.64
6730400	Chain A, Crystal Structure Of An Isolated Apical Domain Of Groel	15702	17.8	53.2	1	4.86
38492781	Chain A, Crystal Structure Of Groel-Groes	55124	5.0	53.2	1	4.87
2624772	Chain A, Crystal Structure Of The Asymmetric Chaperonin Complex GroelGROES(ADP)7	57162	4.8	53.2	1	4.85
11513555	Chain A, Crystal Structure Of The Hexa-Substituted Mutant Of The Molecular Chaperonin Groel Apical	20816	13.5	53.2	1	5.04
11513557	Chain A, Crystal Structure Of The Hexa-Substituted Mutant Of The Molecular Chaperonin Groel Apical	20818	13.5	53.2	1	5.04
20151221	Chain A, Gro-El Fragment (Apical Domain) Comprising Residues 188-379	20692	13.5	53.2	1	4.79
18655582	Chain A, Solution Structure Of Apo Groel By Cryo-Electron Microscopy	57091	4.8	53.2	1	4.81
39654074	Chain A, Structural And Mechanistic Basis For Allostery In The Bacterial Chaperonin Groel; See Rema	57060	4.8	53.2	1	4.90
29726355	Chain A, Structural Basis For Groel-Assisted Protein Folding	57033	4.8	53.2	1	4.84

	From The Crystal Structure Of (Groel-K					
37963076	chaperonin 60 [bacterium DRB3-204]	19493	14.1	53.2	1	4.67
7443844	chaperonin groEL-like protein - Weevil	57007	4.8	53.2	1	4.92
42629991	COG0459: Chaperonin GroEL (HSP60 family) [Haemophilus influenzae R2846]	57763	4.7	53.2	1	4.86
42630824	COG0459: Chaperonin GroEL (HSP60 family) [Haemophilus influenzae R2866]	57518	4.7	53.2	1	4.86
22121898	Cpn60 [Actinobacillus actinomycetemcomitans]	19601	14.1	53.2	1	4.58
2554660	Groel (Hsp60 Class) Fragment (Apical Domain) Comprising Residues 191-376, Mutant With Ala 262 Repl	22016	12.8	53.2	1	5.73
1942200	Groel (Hsp60 Class) Fragment Comprising Residues 191 - 345	16614	16.8	53.2	1	4.82
25988672	GroEL [Brenneria alni]	17257	16.0	53.2	1	5.39
388217	ORF [Yersinia enterocolitica]	19247	14.1	53.2	1	4.70
466577	Yersinia enterocolitica hsp60, cross-reacting protein antigen	57550	4.7	53.2	1	4.88
15228914	calmodulin-binding protein [Arabidopsis thaliana]	96856	0.8	53.1	1	8.07
11498141	DNA gyrase, subunit B (gyrB) [Archaeoglobus fulgidus DSM 4304]	57399	1.4	53.1	1	9.04
23491367	hypothetical protein [Plasmodium yoelii yoelii]	19691	4.2	53.1	1	4.75
13474495	hypothetical protein mll5392 [Mesorhizobium loti MAFF303099]	56296	1.4	53.1	1	6.75
23027223	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [Microbulbifer degradans 2-40]	37323	2.1	53.1	1	5.05
12230403	Probable proline iminopeptidase (PIP) (Prolyl aminopeptidase) (PAP)	36452	2.2	53.1	1	5.81
42524833	proline iminopeptidase [Bdellovibrio bacteriovorus HD100]	36039	2.2	53.1	1	5.86
17987891	PROLINE IMINOPEPTIDASE [Brucella melitensis]	38945	2.1	53.1	1	6.23
23501221	proline iminopeptidase [Brucella suis 1330]	36523	2.2	53.1	1	5.94
21328237	putative cardiolipin synthase [Listeria monocytogenes]	56661	1.4	53.1	1	7.79
29826626	putative erythropoiesis-stimulating protein [Streptomyces avermitilis MA-4680]	35832	2.1	53.1	1	5.87
29831053	putative LuxR-family transcriptional regulator [Streptomyces avermitilis MA-4680]	43360	1.7	53.1	1	7.25
7670023	transcription factor-like protein [Arabidopsis thaliana]	95322	0.8	53.1	1	8.65
23238152	putative RNA-dependent RNA polymerase VP1 [Kadipiro virus]	138459	1.2	53.1	2	8.08
37522961	hypothetical protein glr3392 [Gloeobacter violaceus PCC 7421]	22417	6.1	53.0	1	5.54
14422251	3-oxyacyl-[acyl carrier protein] reductase [Brassica napus]	33120	2.5	53.0	1	9.48
14422263	beta-oxyacyl-[acyl-carrier protein] reductase [Brassica napus]	33650	2.5	53.0	1	9.50
7452557	hypothetical protein F18C5.4 - Caenorhabditis elegans	77991	1.6	52.9	2	9.06
15609992	gorA [Mycobacterium tuberculosis H37Rv]	49914	2.0	52.9	2	5.41
20334298	OmCLC-K [Oreochromis mossambicus]	74786	1.5	52.8	2	8.37
16804182	weakly similar to mannose-6-phosphate isomerase [Listeria monocytogenes EGD-e]	66326	1.7	52.7	1	5.07
410319	with local similarity to fsn (fly) and Ring3; putative	24104	5.1	52.7	2	9.76
24650611	CG3339-PA [Drosophila melanogaster]	521251	0.5	52.7	2	6.11
42490751	agrin [Mus musculus]	198119	0.7	52.6	2	5.81
6754362	insulin receptor-related receptor [Mus musculus]	144653	1.7	52.5	3	6.20
32479251	ferritin GF2 [Crassostrea gigas]	19916	4.1	52.5	1	5.15
33333949	ferritin-like protein [Pinctada fucata]	23564	3.4	52.5	1	5.61
42524205	hypothetical protein predicted by Glimmer/Critica [Bdellovibrio bacteriovorus HD100]	34125	2.3	52.5	1	7.71
1169742	Soma ferritin	20128	4.0	52.5	1	5.16
46097791	hypothetical protein UM05150.1 [Ustilago maydis 521]	200332	0.6	52.4	2	9.47

42526000	methyl-accepting chemotaxis protein [Treponema denticola ATCC 35405]	76539	1.4	52.4	2	5.19
23508722	hypothetical protein [Plasmodium falciparum 3D7]	130240	1.0	52.1	2	9.41
39588960	Hypothetical protein CBG15812 [Caenorhabditis briggsae]	223656	0.7	52.0	2	5.72
27763987	VAB-10A protein [Caenorhabditis elegans]	389314	0.4	52.0	2	5.56
27801756	VAB-10A protein [Caenorhabditis elegans]	389386	0.4	52.0	2	5.55
17511191	VAB-10A protein, Variable ABnormal morphology VAB-10 (vab-10) [Caenorhabditis elegans]	223000	0.7	52.0	2	5.72
42564167	C2 domain-containing protein [Arabidopsis thaliana]	83393	1.4	51.9	1	8.49
15894911	Protein from bacterioferritin family [Clostridium acetobutylicum]	27082	3.8	51.9	1	5.17
42820781	nonribosomal peptide synthetase [Brevibacillus brevis]	567103	0.2	51.8	2	5.10
23510237	erythrocyte membrane protein 1 (PfEMP1) [Plasmodium falciparum 3D7]	404955	0.6	51.8	3	6.06
33599734	putative aldehyde dehydrogenase [Bordetella bronchiseptica RB50]	49943	1.7	51.8	1	5.96
23484385	hypothetical protein [Plasmodium yoelii yoelii]	99075	1.4	51.7	2	9.24
39933239	possible GTP-binding proteins [Rhodopseudomonas palustris CGA009]	37720	2.3	51.7	2	7.74
15228825	pentatricopeptide (PPR) repeat-containing protein [Arabidopsis thaliana]	63968	2.0	51.7	4	6.57
22972523	hypothetical protein [Chloroflexus aurantiacus]	150874	1.4	51.6	2	4.72
13357827	valyl-tRNA synthetase [Ureaplasma urealyticum]	103324	1.5	51.6	2	8.19
23508095	hypothetical protein [Plasmodium falciparum 3D7]	201243	0.9	51.1	2	8.51
25405912	protein F21D18.22 [imported] - Arabidopsis thaliana	297021	0.4	51.1	2	8.10
311664	[Saccharomyces cerevisiae]	77085	1.6	51.0	2	5.75
608567	Bdf1p	76959	1.6	51.0	2	5.82
6648536	moesin [Xenopus laevis]	68092	1.6	51.0	1	5.91
18311295	conserved hypothetical protein [Clostridium perfringens]	34184	2.7	50.9	1	9.94
45549214	CG8779-PA [Drosophila melanogaster]	90688	0.9	50.8	1	5.33
7511976	neuromusculin - fruit fly (Drosophila melanogaster)	113055	0.7	50.8	1	5.24
28493316	transcriptional regulator [Tropheryma whipplei str. Twist]	17958	6.3	50.7	2	6.46
23019737	Protoheme ferro-lyase (ferrochelatase) [Thermobifida fusca]	38882	2.6	50.7	2	5.16
46135873	hypothetical protein FG09452.1 [Gibberella zeae PH-1]	219305	0.7	50.6	2	7.09
45201184	AGR089Cp [Eremothecium gossypii]	156043	1.2	50.5	2	6.06
16127350	AcrB/AcrD/AcrF family protein [Caulobacter crescentus CB15]	109788	1.0	50.4	2	8.90
23509864	hypothetical protein [Plasmodium falciparum 3D7]	35173	3.7	50.3	2	9.88
15672293	ABC transporter ATP binding and permease protein [Lactococcus lactis subsp. lactis]	73979	1.8	50.1	2	5.53
33300661	putative binding protein 7a5 [Homo sapiens]	96578	1.1	35.0	2	6.43
24665073	CG32152-PA [Drosophila melanogaster]	60409	1.7	34.5	2	9.35
555026	core protein	55235	2.4	23.8	2	9.19
46193696	COG0143: Methionyl-tRNA synthetase [Rhodobacter sphaeroides]	64505	1.4	23.5	2	5.36

8.4 Appendix 4: Search results from Conus textile and Conus ventricosus protease treatment.

fraction - number	Accession Number	Protein Description	Mass (Da)	Sequence Coverage	Mascot Protein Score	Number of Peptides	pI
v-a07-e01	18416277	14-3-3 protein GF14 lambda (GRF6) (AFT1) [Arabidopsis thaliana]	27958	5.6	54.5	2	4.77
v-a10-b10	24620453	2MDa_1 protein [Caenorhabditis elegans]	2052136	0.1	48.0	3	4.99
v-a07-c07	129378	60 kDa heat shock protein mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock	60950	4.4	95.1	1	5.83
t-a09-b10	9230281	60S ribosomal protein L2 [Nicotiana tabacum]	18350	4.0	52.1	1	10.87
v-a11-c02	5821952	A Chain A Rotamer Strain As A Determinant Of Protein Structural Specificity	8560	11.8	51.5	1	6.56
v-a07-d02	224305	actin	41542	4.5	92.3	1	5.30
v-a07-c01	113314	ACVS_EMENI N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase (Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase) (ACV synthetase) (ACVS)	422192	0.3	56.0	4	6.01
t-a09-b11	134183	Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)	47053	2.8	52.4	1	5.97
t-a09-b11	3914953	Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)	48735	2.7	52.4	1	6.15
t-a09-b11	21431841	Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase) (Liver copper binding prot	47657	2.8	50.7	1	6.08
v-a07-c07	45187687	ADL186Cp [Eremothecium gossypii]	150320	0.8	56.0	2	6.29
t-a09-b11	45198761	AFR243Cp [Eremothecium gossypii]	49063	2.7	52.4	1	5.58
v-a10-c12	45200825	AGL272Cp [Eremothecium gossypii]	58906	1.8	76.4	2	9.28
v-a10-c12	15889879	AGR_C_4757p [Agrobacterium tumefaciens str. C58]	54590	1.4	54.7	1	6.51
v-a07-c07	21064097	AT16985p [Drosophila melanogaster]	61520	4.3	95.1	1	6.28
v-a07-e01	14335138	AT5g10450/F12B17_200 [Arabidopsis thaliana]	10287	15.6	54.5	2	4.90
v-a10-c12	2493020	ATP synthase alpha chain	54830	1.4	54.7	1	6.31
v-a10-c12	6680748	ATP synthase H+ transporting mitochondrial F1 complex alpha subunit isoform 1 [Mus musculus]	59716	1.3	54.7	1	9.22
v-a10-c12	32766606	Atp5a1-prov protein [Xenopus laevis]	59911	1.3	54.7	1	9.13
v-a07-d02	309090	A-X actin	41667	4.5	74.8	1	5.21
v-a07-d02	6693629	B-actin [Pagrus major]	41813	4.5	89.4	1	5.30
t-a09-b11	10241525	bK3216D2.1.1 (S-adenosylhomocysteine hydrolase (SAHH) isoform 1) [Homo sapiens]	30852	4.2	50.7	1	5.12
v-a07-e01	27376704	bll1593 [Bradyrhizobium japonicum]	88438	1.0	52.4	2	5.18
v-a10-c12	30577817	CAD90949.1 | LssE protein [Legionella pneumophila]	95729	1.5	52.1	2	7.82
v-a07-c07	24641191	CG12101-PA [Drosophila melanogaster]	60771	4.4	95.1	1	5.38
t-a09-b10	17864318	CG1263-PA [Drosophila melanogaster]	27875	2.7	50.5	1	11.15
v-a09-b06	45549186	CG3385-PA [Drosophila melanogaster]	76451	1.7	54.0	2	8.57

v-a10-c12	24658560	CG3612-PA [Drosophila melanogaster]	59384	1.3	54.7	1	9.09
v-a07-e01	17137216	CG5771-PB [Drosophila melanogaster]	24230	5.1	57.6	1	5.53
v-a07-c07	45550132	CG7235-PC [Drosophila melanogaster]	61548	4.3	95.1	1	6.75
v-a10-c12	1943080	Chain A Bovine Mitochondrial F1-Atpase	55199	1.4	54.7	1	8.27
v-a07-d02	2914482	Chain A Complex Of The Second Kunitz Domain Of Tissue Factor Pathway Inhibitor With Porcine Trypsi	23460	3.6	54.4	1	8.26
t-a09-b11	13096481	Chain A Crystal Structure Of Recombinant Rat-Liver D244e Mutant S- Adenosylhomocysteine Hydrolase	47390	2.8	50.7	1	6.08
t-a09-b11	13096485	Chain A Crystal Structure Of S-Adenosylhomocysteine Hydrolase (Adohcyase) Complexed With A Potent	47323	2.8	50.7	1	6.25
v-a07-d02	1942351	Chain A Crystal Structure Of The First Active Autolysate Form Of The Porcine Alpha Trypsin	13284	6.4	54.4	1	7.83
v-a10-c12	6729934	Chain A Rat Liver F1-Atpase	55247	1.4	54.7	1	8.28
t-a09-b11	4139571	Chain A Rat Liver S-Adenosylhomocystein Hydrolase	47376	2.8	50.7	1	6.08
v-a07-e01	5821952	Chain A Rotamer Strain As A Determinant Of Protein Structural Specificity	8560	80.3	276.3	12	6.56
v-a07-e01	5107695	Chain A Solution Structure Of The Designed Hydrophobic Core Mutant Of Ubiquitin 1d7	8580	56.6	251.9	11	6.56
t-a09-b11	4929854	Chain A Structure Of Human Placental S-Adenosylhomocysteine Hydrolase: Determination Of A 30 Selen	47364	2.8	50.7	1	5.93
v-a07-e01	13786827	Chain A Structure Of Recombinant Human Ubiquitin In Aot Reverse Micelles	8540	88.2	361.2	15	5.77
v-a07-d02	494360	Chain A Trypsin (E.C.3.4.21.4) Complexed With Inhibitor From Bitter Gourd	23458	3.6	54.4	1	8.60
v-a07-e01	42543202	Chain A X-Ray Structure Of The Small G Protein Rab11a In Complex With Gdp	21495	5.8	57.6	1	6.50
v-a07-e01	42543204	Chain A X-Ray Structure Of The Small G Protein Rab11a In Complex With Gtpgammas	21480	5.8	57.6	1	6.50
v-a07-d02	999627	Chain B Porcine E-Trypsin (E.C.3.4.21.4)	8814	9.8	54.4	1	6.67
v-a07-d02	3318722	Chain E Leech-Derived Tryptase InhibitorTRYPSIN COMPLEX	23457	3.6	54.4	1	8.26
v-a07-c07	306890	chaperonin (HSP60)	60986	4.4	95.1	1	5.70
v-a07-c07	21805770	chaperonin Cpn60 [Danio rerio]	33309	8.1	95.1	1	9.13
v-a07-e01	21232833	choline dehydrogenase [Xanthomonas campestris pv. campestris str. ATCC 33913]	61406	1.8	56.5	3	6.26
v-a10-c12	23000523	COG0056: F0F1-type ATP synthase alpha subunit [Magnetococcus sp. MC-1]	54600	1.4	53.0	1	5.27
v-a10-c12	23015477	COG0056: F0F1-type ATP synthase alpha subunit [Magnetospirillum magnetotacticum]	54831	1.4	54.7	1	6.31
t-a09-b10	23051982	COG0090: Ribosomal protein L2 [Methanosarcina barkeri]	25713	2.9	52.1	1	10.43
t-a09-b11	23103453	COG0499: S-adenosylhomocysteine hydrolase [Azotobacter vinelandii]	72430	1.8	52.4	1	9.26
t-a11-c10	41726013	COG0583: Transcriptional regulator [Dechloromonas aromatica RCB]	32829	2.3	52.9	1	5.52

v-a07-c01	22956766	COG1009: NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter MnhA subunit [Rhodobacter sphaeroides]	103735	0.9	50.2	4	9.28
v-a10-c12	45915510	COG1171: Threonine dehydratase [Mesorhizobium sp. BNC1]	45821	1.9	51.7	2	6.99
v-a10-c01	22997572	COG2925: Exonuclease I [Xylella fastidiosa Ann-1]	56045	2.7	52.4	2	6.16
v-a07-e01	46363489	COG5001: Predicted signal transduction protein containing a membrane domain an EAL and a GGDEF dom	76585	2.8	68.2	2	5.64
t-a09-c08	46113353	COG5001: Predicted signal transduction protein containing a membrane domain an EAL and a GGDEF domain [Exiguobacterium sp. 255-15]	104409	0.9	55.1	1	5.19
t-a11-c02	**12619439**	**conotoxin scaffold III/IV precursor [*Conus textile*]**	**7919**	**22.9**	**58.4**	**1**	**5.07**
v-a07-c07	**12619397**	**conotoxin scaffold III/IV precursor [*Conus ventricosus*]**	**7393**	**25.4**	**51.1**	**2**	**4.51**
v-a07-c01	**12619623**	**conotoxin scaffold VI/VII precursor [*Conus ventricosus*]**	**8096**	**26.3**	**123.7**	**3**	**4.76**
v-a09-c11	**12619613**	**conotoxin scaffold VI/VII precursor [*Conus ventricosus*]**	**7883**	**13.9**	**50.9**	**1**	**7.52**
v-a09-c11	**12619615**	**conotoxin scaffold VI/VII precursor [*Conus ventricosus*]**	**7882**	**13.9**	**50.9**	**1**	**8.16**
v-a10-c12	19745428	conserved hypothetical protein [Streptococcus pyogenes MGAS8232]	41667	2.2	23.7	2	6.23
t-a09-b10	46101825	conserved hypothetical protein [Ustilago maydis 521]	10666	7.0	52.1	1	12.25
t-a07-c11	21218683	conserved hypothetical protein SCJ2.04c [Streptomyces coelicolor A3(2)]	48605	3.1	53.7	2	9.19
v-a07-d02	6230866	cytoplasmic actin [Oikopleura longicauda]	41672	4.5	92.3	1	5.23
v-a10-c12	17507955	defective (57.8 kD) (1N245) [Caenorhabditis elegans]	57752	1.3	54.7	1	8.98
v-a07-e01	2134996	desmoplakin I - human	309797	0.6	80.1	2	6.36
t-a09-c08	4758200	desmoplakin; desmoplakin (DPI DPII) [Homo sapiens]	331571	1.2	99.7	4	6.44
v-a07-e01	4758200	desmoplakin; desmoplakin (DPI DPII) [Homo sapiens]	331571	0.6	80.1	2	6.44
v-a07-e01	34764046	dGTP triphosphohydrolase [Fusobacterium nucleatum subsp. vincentii ATCC 49256]	64179	2.0	54.8	2	8.52
v-a07-e01	86248	dolichyl-diphosphooligosaccharide-protein glycotransferase (EC 2.4.1.119) glycosylation site-binding chain precursor - chicken	56929	1.6	57.0	3	4.84
v-a10-c12	32307378	ecdysteroid UDP-glucosyltransferase [Spodoptera frugiperda MNPV]	60598	1.5	38.2	2	6.61
v-a10-c01	31204473	ENSANGP00000001527 [Anopheles gambiae]	244959	0.8	53.8	3	5.48
v-a10-c12	31210101	ENSANGP00000009989 [Anopheles gambiae]	59256	1.3	54.7	1	9.14
t-a09-b10	31214173	ENSANGP00000010416 [Anopheles gambiae]	28967	2.7	49.5	1	10.91
v-a07-c07	31231072	ENSANGP00000014839 [Anopheles gambiae]	61924	4.3	95.1	1	5.56
v-a10-c12	31208079	ENSANGP00000020317 [Anopheles gambiae]	46810	2.9	57.2	2	9.36
v-a10-c12	31207617	ENSANGP00000020421 [Anopheles gambiae]	34353	3.0	53.5	1	9.14

v-a07-e01	31209781	ENSANGP00000024287 [Anopheles gambiae]	24332	5.1	57.6	1	5.73
v-a10-c01	15838616	exodeoxyribonuclease I [Xylella fastidiosa 9a5c]	56822	2.6	52.4	2	5.93
v-a10-c01	28198698	exodeoxyribonuclease I [Xylella fastidiosa Temecula1]	56849	2.6	52.4	2	6.16
v-a07-e01	38077190	expressed sequence AI507495 [Mus musculus]	56715	6.9	188.3	6	7.08
v-a07-e01	1405561	FSGTP1 [Fagus sylvatica]	23919	5.1	57.6	1	9.23
v-a10-c12	11610624	GABA-A epsilon subunit splice variant [Rattus norvegicus]	45976	1.8	58.2	1	9.04
v-a10-c12	11610626	GABA-A epsilon subunit splice variant [Rattus norvegicus]	45893	1.8	58.2	1	9.00
v-a10-c12	11610622	GABA-A epsilon subunit splice variant [Rattus norvegicus]	94974	0.9	58.2	1	4.38
v-a10-c12	2735329	GABA-A receptor epsilon subunit [Rattus norvegicus]	45436	1.8	58.2	1	9.26
v-a10-c12	12831207	gamma-aminobutyric acid A receptor epsilon; gamma-aminobutyric acid (GABA-A) receptor subunit epsilon; gamma-aminobutyric acid (GABA-A) receptor subunit epsilon [Rattus norvegicus]	113220	0.7	58.2	1	4.41
v-a10-c12	13959374	Gamma-aminobutyric-acid receptor epsilon subunit precursor (GABA(A) receptor)	59180	1.4	58.2	1	9.48
t-a09-c08	29650759	gamma-catenin [Homo sapiens]	81665	4.7	83.2	3	5.75
v-a07-e01	11558850	11558850|emb|CAC17961.1| vimentin [Scyliorhinus stellaris]	51948	2.9	44.0	3	5.04
v-a07-e01	14590469	14590469|ref|NP_142537.1| pyruvate kinase [Pyrococcus horikoshii]	53141	4.8	63.5	3	8.78
v-a07-e01	46127519	46127519|ref|XP_388313.1| hypothetical protein FG08137.1 [Gibberella zeae PH-1]	32158	3.8	57.8	2	5.21
t-a09-b10	29246632	GLP_13_32668_33423 [Giardia lamblia ATCC 50803]	27020	2.8	52.1	1	11.15
v-a07-e01	45382295	glycosylation site-binding protein [Gallus gallus]	56856	1.6	57.0	3	4.84
v-a07-e01	7438432	GTP-binding protein - garden pea	24088	5.1	57.6	1	5.65
v-a07-e01	21555752	guanine nucleotide regulatory protein putative [Arabidopsis thaliana]	24039	5.1	57.6	1	5.59
v-a10-c12	28630328	H+-transporting ATP synthase alpha subunit isoform 1 [Branchiostoma lanceolatum]	42659	1.8	54.7	1	8.92
v-a07-c07	31981679	heat shock protein 1 (chaperonin); heat shock protein 60 kDa; heat shock 60kDa protein 1 (chaperon	60918	4.4	95.1	1	5.67
v-a07-c07	11560024	heat shock protein 60 (liver); heat shock 60kD protein 1 (chaperonin) [Rattus norvegicus]	60927	4.4	95.1	1	5.91
v-a07-e01	16801015	highly similar to carbamoyl-phosphate synthetase (catalytic subunit) [Listeria innocua]	117788	0.9	52.7	2	4.80
v-a09-e01	17565064	histone 3.3 (15.3 kD) (5N140) [Caenorhabditis elegans]	15340	13.2	52.7	2	11.40
v-a09-e01	122007	Histone H2A	15823	13.4	52.8	2	10.66
v-a09-e01	15238430	histone H3 [Arabidopsis thaliana]	15582	12.9	52.7	2	11.26
v-a07-d02	1197519	histone H3 [Narcissus pseudonarcissus]	14396	7.8	92.8	3	11.12
v-a09-b06	1197519	histone H3 [Narcissus pseudonarcissus]	14396	5.5	67.1	2	11.12
v-a09-e01	223582	histone H4	11230	29.4	109.5	3	11.20
v-a09-b06	1708110	Histone H4	11450	17.5	80.0	2	11.48

v-a09-e01	1708110	Histone H4	11450	17.5	75.9	2	11.48
t-a11-c02	23477111	histone H4 [Colletotrichum sp.]	6814	16.7	56.3	1	10.67
v-a07-d02	23477111	histone H4 [Colletotrichum sp.]	6814	16.7	54.5	1	10.67
t-a11-c02	23476928	histone H4 [Glomerella acutata]	7256	15.6	56.3	1	10.94
v-a07-d02	23476928	histone H4 [Glomerella acutata]	7256	15.6	54.5	1	10.94
t-a11-c02	23476924	histone H4 [Glomerella cingulata]	5696	20.0	56.3	1	10.96
v-a07-d02	23476924	histone H4 [Glomerella cingulata]	5696	20.0	54.5	1	10.96
t-a11-c02	23476926	histone H4 [Glomerella graminicola]	3812	29.4	56.3	1	11.36
v-a07-d02	23476926	histone H4 [Glomerella graminicola]	3812	29.4	54.5	1	11.36
v-a10-c12	30585339	Homo sapiens ATP synthase H+ transporting mitochondrial F1 complex alpha subunit isoform 1 car	59827	1.3	54.7	1	9.16
v-a07-e01	30584069	Homo sapiens RAB11A member RAS oncogene family [synthetic construct]	24492	5.1	57.6	1	6.12
t-a09-b11	30584089	Homo sapiens S-adenosylhomocysteine hydrolase [synthetic construct]	47798	2.8	50.7	1	5.92
v-a07-c07	27735378	Hspd1 protein [Xenopus laevis]	59616	4.5	95.1	1	6.62
v-a07-c01	37533392	hypothetical protein [Oryza sativa (japonica cultivar-group)]	41979	2.9	50.7	4	7.21
v-a10-b10	40741499	hypothetical protein AN8655.2 [Aspergillus nidulans FGSC A4]	81447	1.1	54.3	2	6.39
v-a07-e01	39595276	Hypothetical protein CBG03904 [Caenorhabditis briggsae]	23400	5.2	57.6	1	6.44
t-a09-b10	39585332	Hypothetical protein CBG05588 [Caenorhabditis briggsae]	28173	2.7	50.5	1	11.07
v-a10-c12	39580722	Hypothetical protein CBG08717 [Caenorhabditis briggsae]	57667	1.3	54.7	1	8.98
v-a10-b10	39752855	Hypothetical protein F12F3.2a [Caenorhabditis elegans]	700957	0.3	50.7	3	5.72
v-a10-c12	31441843	Hypothetical protein H28O16.1c [Caenorhabditis elegans]	54121	1.4	54.7	1	9.10
v-a07-d02	38103539	hypothetical protein MG03982.4 [Magnaporthe grisea 70-15]	39700	4.8	64.1	3	5.62
v-a07-c01	38107610	hypothetical protein MG09507.4 [Magnaporthe grisea 70-15]	76989	4.9	53.6	5	9.52
t-a09-b10	41152307	hypothetical protein MGC73105 [Danio rerio]	28049	2.7	52.1	1	10.89
v-a07-c01	8745001	hypothetical protein P1295.15 [Leishmania major]	303147	0.3	51.5	4	6.33
v-a10-b10	46101547	hypothetical protein UM05835.1 [Ustilago maydis 521]	136342	0.7	54.3	2	8.74
v-a07-c07	34875806	hypothetical protein XP_212759 [Rattus norvegicus]	60967	4.4	95.1	1	5.91
v-a07-e01	7513694	IgG Fc binding protein - mouse (fragment)	109137	0.9	56.4	2	4.71
t-a09-c08	41529837	junction plakoglobin; gamma-catenin (plakoglobin) [Rattus norvegicus]	81749	4.7	83.2	3	5.75
v-a07-c01	34784398	Krt1-15 protein [Mus musculus]	49463	7.5	201.1	5	4.79
v-a07-e01	39850082	Krt2-4 protein [Mus musculus]	57817	10.4	278.6	9	8.09
v-a07-e01	38303961	Krt5 protein [Rattus norvegicus]	61788	14.4	374.6	13	7.60
t-a09-c08	38303961	Krt5 protein [Rattus norvegicus]	61788	8.2	288.2	7	7.60
v-a07-e01	40787691	LOC395048 protein [Xenopus tropicalis]	56745	1.6	55.3	3	4.73
t-a09-b10	21228229	LSU ribosomal protein L2P [Methanosarcina	25729	2.9	52.1	1	10.14

		mazei Goe1]					
v-a10-b10	22966565	Methyl-accepting chemotaxis protein [Rhodospirillum rubrum]	46733	1.8	54.3	2	5.44
v-a07-c07	28436902	MGC53106 protein [Xenopus laevis]	50499	5.3	95.1	1	7.57
v-a07-c07	40647591	mitochondrial 60 kDa heat shock protein [Anemonia viridis]	62768	4.3	95.1	1	5.30
v-a10-c12	14009437	mitochondrial ATP synthase alpha-subunit [Cyprinus carpio]	59528	1.3	54.7	1	9.33
v-a07-d02	28336	mutant beta-actin (beta'-actin) [Homo sapiens]	41786	4.5	74.8	1	5.22
v-a09-b06	790600	nervy	76512	1.7	54.0	2	8.83
v-a10-c12	17227664	NP_484212.1| alpha-amylase [Nostoc sp. PCC 7120]	70850	1.3	51.7	2	5.16
v-a10-c12	29828972	NP_823606.1| putative two-component system sensor kinase [Streptomyces avermitilis MA-4680]	61446	1.7	52.7	2	5.44
v-a11-c02	9634344	ORF123 v-ubiquitin [Spodoptera exigua nucleopolyhedrovirus]	9358	11.3	51.5	1	8.04
v-a10-c12	38345636	OSJNBb0048E02.6 [Oryza sativa (japonica cultivar-group)]	38421	2.3	51.7	2	10.33
v-a07-e01	28436918	P4hb protein [Xenopus laevis]	57980	1.5	55.3	3	4.81
t-a09-c08	130257	PLAK_HUMAN Junction plakoglobin (Desmoplakin III)	81578	4.7	83.2	3	5.95
t-a09-c08	762885	Plakoglobin	81583	4.7	83.2	3	5.88
t-a09-c08	20336613	plakoglobin [Bos taurus]	81769	4.7	83.2	3	5.75
v-a11-c02	10121776	polyprotein [bovine viral diarrhea virus type 2]	71690	3.3	64.7	2	9.17
v-a07-e01	1101011	polyubiquitin	87122	4.8	105.5	6	6.36
v-a07-e01	3789942	polyubiquitin [Saccharum hybrid cultivar H32-8560]	42642	13.6	216.5	11	8.43
v-a10-c12	15828596	predicted coding region [Mycoplasma pulmonis]	64970	1.5	58.2	1	9.12
v-a07-c07	11386856	Probable 60 kDa heat shock protein homolog 2 mitochondrial precursor (Hsp60) (60 kDa chaperonin) (	61130	4.4	95.1	1	6.28
v-a10-c12	15966789	PROBABLE ATP SYNTHASE ALPHA CHAIN PROTEIN [Sinorhizobium meliloti 1021]	54647	1.4	54.7	1	6.30
v-a07-e01	18376163	probable GTP-binding protein Drab11 [Neurospora crassa]	23704	5.1	57.6	1	5.61
v-a10-b10	15240463	protein transport protein-related [Arabidopsis thaliana]	135746	1.0	53.6	2	5.44
t-a09-b11	19112372	putative adenosylhomocysteinase [Schizosaccharomyces pombe]	47353	2.8	52.4	1	5.61
v-a07-e01	46359908	putative GTP-binding protein [Oryza sativa (japonica cultivar-group)]	23987	5.0	57.6	1	6.13
v-a10-c12	39933255	putative H+-transporting ATP synthase alpha chain. [Rhodopseudomonas palustris CGA009]	55143	1.4	54.7	1	6.14
t-a11-c10	16264227	putative L-xylulose kinase protein [Sinorhizobium meliloti 1021]	54393	2.4	52.5	2	5.46
t-a09-b11	6320882	Putative S-adenosyl-L-homocysteine hydrolase with a probable role in S-adenosylhomocysteine catabol	49094	2.7	52.4	1	5.83
v-a11-c02	22549552	putative ubiquitin [Mamestra configurata	11227	9.0	49.8	1	8.85

		nucleopolyhedrovirus B]					
v-a07-e01	17507543	RAB family member (23.4 kD) (rab-11.1) [Caenorhabditis elegans]	23415	5.2	57.6	1	6.44
v-a07-e01	37788825	Rab11-1a [Limulus polyphemus]	24358	5.1	57.6	1	5.70
t-a09-b10	22758874	ribosomal protein L [Argopecten irradians]	28178	2.7	52.1	1	10.82
t-a09-b10	12054507	ribosomal protein L2 [Glycine max]	27996	2.7	52.1	1	10.45
t-a09-b10	32400873	ribosomal protein L2 [Triticum aestivum]	9558	7.9	52.1	1	11.68
t-a09-b10	20089945	ribosomal protein L2p [Methanosarcina acetivorans str. C2A]	25675	2.9	52.1	1	10.28
t-a09-b10	5690416	ribosomal protein L8 [Anopheles gambiae]	28715	2.7	50.5	1	11.07
t-a09-b10	3047007	ribosomal protein L8 [Aplysia californica]	5747	13.2	52.1	1	11.89
t-a09-b10	20069091	ribosomal protein L8 [Aplysia californica]	9109	8.5	52.1	1	11.29
t-a09-b10	28200282	ribosomal protein L8 [Branchiostoma belcheri tsingtaunese]	8543	9.1	52.1	1	11.78
t-a09-b10	45384773	ribosomal protein L8 [Chinchilla lanigera]	16284	4.6	52.1	1	10.90
t-a09-b10	40643032	ribosomal protein L8 [Crassostrea gigas]	10875	7.0	52.1	1	11.58
t-a09-b10	15082586	Ribosomal protein L8 [Homo sapiens]	27993	2.7	52.1	1	11.03
t-a09-b10	15293881	ribosomal protein L8 [Ictalurus punctatus]	28127	2.7	50.5	1	10.97
t-a09-b10	16566710	ribosomal protein L8 [Spodoptera frugiperda]	27901	2.7	50.5	1	10.76
t-a09-b10	32880093	ribosomal protein L8 [synthetic construct]	28106	2.7	52.1	1	11.03
t-a09-b10	4506663	ribosomal protein L8; 60S ribosomal protein L8 [Homo sapiens]	28007	2.7	52.1	1	11.03
t-a09-b10	17557310	ribosomal Protein Large subunit (28.2 kD) (rpl-2) [Caenorhabditis elegans]	28186	2.7	50.5	1	11.07
v-a07-e01	20903595	RIKEN cDNA 2310001L23 [Mus musculus]	62806	8.4	279.1	11	8.68
t-a09-c08	20903595	RIKEN cDNA 2310001L23 [Mus musculus]	62806	6.6	206.5	8	8.68
v-a07-c01	20903595	RIKEN cDNA 2310001L23 [Mus musculus]	62806	5.1	109.8	3	8.68
v-a10-c04	38091800	RIKEN cDNA 2310058N18 [Mus musculus]	48206	4.0	59.9	2	4.78
v-a07-e01	19526922	RIKEN cDNA 4631426H08 [Mus musculus]	48891	4.0	81.6	3	4.99
v-a07-e01	25056305	RIKEN cDNA 4733401L19 [Mus musculus]	50315	3.5	62.2	2	5.19
t-a09-b10	38014625	RPL8 protein [Homo sapiens]	22886	3.3	52.1	1	10.56
t-a09-b11	178277	S-adenosylhomocysteine hydrolase	47684	2.8	50.7	1	6.03
v-a07-c07	33636453	SD06594p [Drosophila melanogaster]	60758	4.4	95.1	1	5.38
t-a09-b10	27685597	similar to 60S ribosomal protein L8 [Rattus norvegicus]	28139	2.7	52.1	1	10.79
v-a07-e01	38086180	similar to A430096B05Rik protein [Mus musculus]	156503	0.6	56.4	2	5.66
t-a09-b11	34852949	similar to Ahcy protein [Rattus norvegicus]	51886	2.5	52.4	1	5.65
v-a10-c12	13938339	Similar to ATP synthase H+ transporting mitochondrial F1 complex alpha subunit isoform 1 cardi	44362	1.7	54.7	1	9.42
v-a10-c12	24660110	Similar to ATP synthase H+ transporting mitochondrial F1 complex alpha subunit isoform 1 cardi	60748	1.2	54.7	1	9.14
v-a11-c02	34854675	similar to caspase 3 apoptosis related cysteine protease [Rattus norvegicus]	36680	2.8	61.3	2	8.61
v-a07-d02	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	6.9	262.0	7	10.58
v-a09-b06	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	4.9	228.1	5	10.58
v-a09-e01	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	6.8	211.9	7	10.58
t-a10-e01	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	5.0	155.0	6	10.58
t-a11-c02	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	2.1	80.4	2	10.58

v-a09-c11	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	1.8	64.2	2	10.58
t-a09-c08	34876364	similar to CG31613-PA [Rattus norvegicus]	101471	1.8	54.9	2	10.58
v-a07-e01	34875216	similar to Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen) [Rattus norvegicus]	327589	0.6	80.1	2	6.45
t-a09-b10	28829974	similar to Dictyostelium discoideum (Slime mold). 60S ribosomal protein L2	27503	2.7	52.1	1	10.69
v-a07-e01	34855405	similar to Fc fragment of IgG binding protein; IgG Fc binding protein [Rattus norvegicus]	583032	0.2	56.4	2	5.40
v-a09-b06	34876396	similar to H3 histone family member I [Rattus norvegicus]	33259	11.7	115.5	3	11.00
v-a07-e01	38079884	similar to ribosomal protein S27a [Mus musculus]	13110	11.0	52.3	2	8.92
v-a07-e01	34879387	similar to ribosomal protein S27a [Rattus norvegicus]	17897	14.2	88.1	3	9.50
v-a07-e01	34873655	similar to RIKEN cDNA 4631426H08 [Rattus norvegicus]	79826	2.5	81.6	3	4.74
t-a09-b11	20861044	similar to S-adenosylhomocysteine hydrolase [Mus musculus]	26335	5.0	52.4	1	5.96
v-a07-e01	7108528	small GTPase [Mus musculus]	24387	5.1	57.6	1	6.23
t-a11-c10	17987814	TETRATRICOPEPTIDE REPEAT FAMILY PROTEIN [Brucella melitensis]	66415	1.2	52.9	1	5.75
v-a07-e01	3789940	tetra-ubiquitin [Saccharum hybrid cultivar H32-8560]	34202	19.3	226.3	10	6.17
t-a11-c10	23501306	TPR domain protein [Brucella suis 1330]	69274	1.1	52.9	1	6.23
v-a07-d02	136429	Trypsin precursor	24394	3.5	54.4	1	7.00
v-a07-e01	10719701	ubiquitin [Scyliorhinus torazame]	19471	40.5	322.9	15	9.35
v-a07-e01	28194281	ubiquitin extension protein [Heterodera glycines]	11762	32.4	171.3	8	9.65
v-a11-c02	33346945	ubiquitin/actin fusion protein [Gymnochlora stellata]	49352	4.5	78.8	3	5.03
v-a07-e01	33346927	ubiquitin/actin fusion protein 1 [Bigelowiella natans]	26376	20.9	248.7	10	4.83
v-a11-c02	33346927	ubiquitin/actin fusion protein 1 [Bigelowiella natans]	26376	8.4	78.8	3	4.83
v-a07-e01	33346933	ubiquitin/actin fusion protein 1 [Lotharella amoeboformis]	25960	21.3	248.7	10	5.01
v-a11-c02	33346933	ubiquitin/actin fusion protein 1 [Lotharella amoeboformis]	25960	8.5	78.8	3	5.01
v-a07-e01	33346939	ubiquitin/actin fusion protein 1 [Lotharella globosa]	26018	21.1	248.7	10	5.07
v-a11-c02	33346939	ubiquitin/actin fusion protein 1 [Lotharella globosa]	26018	8.4	78.8	3	5.07
v-a07-e01	33304714	ubiquitin/actin fusion protein 2 [Bigelowiella natans]	51097	10.9	247.0	10	5.31
v-a11-c02	33304714	ubiquitin/actin fusion protein 2 [Bigelowiella natans]	51097	4.4	78.8	3	5.31
v-a11-c02	33346935	ubiquitin/actin fusion protein 2 [Lotharella amoeboformis]	25974	8.5	78.8	3	5.01
v-a07-e01	33346943	ubiquitin/actin fusion protein 3 [Lotharella globosa]	26275	21.0	248.7	10	4.90

v-a11-c02	33346943	ubiquitin/actin fusion protein 3 [Lotharella globosa]	26275	8.4	78.8	3	4.90
v-a07-e01	29294722	Unknown (protein for MGC:55760) [Danio rerio]	24565	5.0	57.6	1	5.64
v-a07-e01	41056251	Unknown (protein for MGC:63565) [Danio rerio]	24103	5.1	57.6	1	6.22
v-a07-d02	45872608	Unknown (protein for MGC:76228) [Xenopus tropicalis]	41726	4.5	74.8	1	5.29
t-a09-c08	27364239	Unknown [Vibrio vulnificus CMCP6]	27405	3.3	51.9	1	11.24
v-a07-e01	21751806	unnamed protein product [Homo sapiens]	44670	9.0	188.3	6	5.07
v-a07-c01	21751806	unnamed protein product [Homo sapiens]	44670	6.0	87.6	5	5.07
v-a10-c12	38422323	unnamed protein product [Magnetospirillum gryphiswaldense]	54610	1.4	54.7	1	6.10
v-a07-c07	51452	unnamed protein product [Mus musculus]	58833	4.5	95.1	1	5.48
v-a10-c04	12845122	unnamed protein product [Mus musculus]	33861	5.9	59.9	2	4.67
v-a10-c04	26343627	unnamed protein product [Mus musculus]	47111	4.1	59.9	2	4.75
v-a07-e01	12843679	unnamed protein product [Mus musculus]	55757	3.7	53.4	2	6.19
t-a09-b11	26344433	unnamed protein product [Mus musculus]	36017	3.7	52.4	1	6.08
v-a07-c07	1334284	unnamed protein product [Rattus norvegicus]	57890	4.6	95.1	1	5.35
t-a09-b10	5063	unnamed protein product [Schizosaccharomyces pombe]	27099	2.8	50.5	1	10.86
v-a07-e01	28277305	Vim4 protein [Xenopus laevis]	53507	3.5	64.7	3	5.12
v-a07-e01	32450567	Vim4 protein [Xenopus laevis]	53572	3.5	64.7	3	5.12
v-a07-e01	138532	Vimentin 4	53464	3.5	64.7	3	5.08
v-a07-e01	138531	Vimentins 1 and 2	52812	3.5	64.7	3	5.16
v-a11-c02	20070031	v-ubiquitin [Mamestra configurata nucleopolyhedrovirus A]	11402	9.0	51.5	1	9.89
v-a07-e01	25463443	YPT7 protein homolog [imported] - yeast (Candida albicans)	32027	3.8	57.6	1	7.00
v-a10-c12	22988880	ZP_00033940.1| COG1123: ATPase components of various ABC-type transport systems contain duplicated ATPase [Burkholderia fungorum]	68123	1.3	51.7	2	9.36
v-a10-c12	23051406	ZP_00078165.1| COG1060: Thiamine biosynthesis enzyme ThiH and related uncharacterised enzymes [Methanosarcina barkeri]	36307	2.5	35.5	2	5.96

9 References

Alonso, D., Khalil, Z., Satkunanthan, N. and Livett, B. G. (2003). Drugs from the sea: conotoxins as drug leads for neuropathic pain and other neurological conditions. *Mini Rev Med Chem* **3**, 785-7.

Atanassoff, P. G., Hartmannsgruber, M. W., Thrasher, J., Wermeling, D., Longton, W., Gaeta, R., Singh, T., Mayo, M., McGuire, D. and Luther, R. R. (2000). Ziconotide, a new N-type calcium channel blocker, administered intrathecally for acute postoperative pain. *Reg Anesth Pain Med* **25**, 274-8.

Baldwin, M. A. (2004). Protein Identification by Mass Spectrometry: Issues to be Considered. *Mol.Cell Proteomics* **3**, 1-9.

Banoub, J., Cohen, A., Mansour, A. and Thibault, P. (2004). Characterization and de novo sequencing of Atlantic salmon vitellogenin protein by electrospray tandem and matrix-assisted laser desorption/ionization mass spectrometry. *European Journal of Mass Spectrometry* **10**, 121-134.

Benvenuti, S., Cramer, R., Bruce, J., Waterfield, M. D. and Jat, P. S. (2002). Identification of novel candidates for replicative senescence by functional proteomics. *Oncogene* **21**, 4403-4413.

Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. and Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**, 317-39.

Blanchfield, J. T., Dutton, J. L., Hogg, R. C., Gallagher, O. P., Craik, D. J., Jones, A., Adams, D. J., Lewis, R. J., Alewood, P. F. and Toth, I. (2003). Synthesis, Structure Elucidation, in Vitro Biological Activity, Toxicity, and Caco-2 Cell Permeability of Lipophilic Analogues of alpha-Conotoxin MII. *J.Med.Chem.* **46**, 1266-1272.

Bowersox, S. S. and Luther, R. (1998). Pharmacotherapeutic potential of omega-conotoxin MVIIA (SNX-111), an N-type neuronal calcium channel blocker found in the venom of Conus magus. *Toxicon* **36**, 1651-8.

Boyle, J. G. and Whitehouse, C. M. (1992). Time-of-flight mass spectrometry with an electrospray ion beam. *Anal Chem* **64**, 2084-9.

Burlet, O., Orkiszewski, R. S., Ballard, K. D. and Gaskell, S. J. (1992). Charge promotion of low-energy fragmentations of peptide ions. *Rapid Communications in Mass Spectrometry* **6**, 658.

Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E. and Karger, B. L. (2005). High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* **77**, 2323-31.

Clauser, K. R., Baker, P. and Burlingame, A. L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **71**, 2871-82.

Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z. and Fainzilber, M. (2001). Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol* **18**, 120-31.

Conticello, S. G., Pilpel, Y., Glusman, G. and Fainzilber, M. (2000). Position-specific codon conservation in hypervariable gene families. *Trends Genet* **16**, 57-9.

Cornish, T. J. and Cotter, R. J. (1993). A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **7**, 1037-40.

Cotter, R. (1984). Lasers and mass spectrometry. *Anal Biochem* **56**, 485A-504A.

Cotter, R. J. (1989). Time-of-flight mass spectrometry: an increasing role in the life sciences. *Biomed Environ Mass Spectrom* **18**, 513-32.

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. et al. (1997). 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.* **272**, 4689-4698.

Cramer, R. and Burlingame, A. L. (2000). Employing target modifications for the investigation of liquid infrared matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* **14**, 53-60.

Cramer, R. and Corless, S. (2001). The nature of collision-induced dissociation processes of doubly protonated peptides: comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **15**, 2058-2066.

Cramer, R. and Corless, S. (2005). Liquid ultraviolet matrix-assisted laser desorption/ionization - mass spectrometry for automated proteomic analysis. *Proteomics.*

Cramer, R., Richter, W., Stimson, E. and Burlingame, A. (1998). Analysis of phospho- and glycopolypeptides with infrared matrix-assisted laser desorption and ionization. *Anal Chem* **70**, 4939-4944.

Cruz, L. J., Gray, W. R., Olivera, B. M., Zeikus, R. D., Kerr, L., Yoshikami, D. and Moczydlowski, E. (1985). Conus-Geographus Toxins That Discriminate Between

Neuronal and Muscle Sodium-Channels. *Journal of Biological Chemistry* **260**, 9280-9288.

Cruz, L. J. and White, J. (1995). Clinical toxicoogy of Conus snail stings.: CRC.

Csonka, I. P., Paizs, B. and Suhai, S. (2004). Modeling of the gas-phase ion chemistry of protonated arginine. *Journal of Mass Spectrometry* **39**, 1025-1035.

Cutillas, P., Burlingame, A. and Unwin, R. (2004). Proteomic strategies and their application in studies of renal function. *News Physiol Sci.* **19**, 114-119.

Dreisewerd, K. (2003). The desorption process in MALDI. *Chem.Rev.* **103**, 395-426.

Duda, T. F., Jr. and 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**, 6820-3.

E, J., C, H., P, H. and B, L. (2001). Determination of paralytic shellfish poisoning toxins by high-performance ion-exchange chromatography. *J Chromatogr A* **929**, 43-49.

Ehring, H., Karas, M. and Hillenkamp, F. (1992). Role of photoionization and photochemistry in ionization processes of organic molecules and relevance for matrix-assisted laser desorption uioniztion mass spectrometry. *Org. Mass spectrom.* **27**, 472-480.

Escoubas, P., Diochot, S. and Corzo, G. (2000). Structure and pharmacology of spider venom neurotoxins. *Biochimie* **82**, 893-907.

Evans, R. M. (1998). Vimentin: the conundrum of the intermediate filament gene family. *Bioessays* **20**, 79-86.

Fainzilber, M., Gordon, D., Hasson, A., Spira, M. E. and Zlotkin, E. (1991). Mollusc-specific toxins from the venom of Conus textile neovicarius. *Eur J Biochem* **202**, 589-95.

Fainzilber, M., Lodder, J. C., Kits, K. S., Kofman, O., Vinnitsky, I., Van Rietschoten, J., Zlotkin, E. and Gordon, D. (1995a). A new conotoxin affecting sodium current inactivation interacts with the delta-conotoxin receptor site. *J Biol Chem* **270**, 1123-9.

Fainzilber, M., Nakamura, T., Gaathon, A., Lodder, J. C., Kits, K. S., Burlingame, A. L. and Zlotkin, E. (1995b). A new cysteine framework in sodium channel blocking conotoxins. *Biochemistry* **34**, 8649-56.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71.

Fitzgerald, M. C., Parr, G. R. and Smith, L. M. (1993). Basic matrices for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides. *Anal Chem* **65**, 3204-11.

Floyd, P. D., Li, L., Moroz, T. P. and Sweedler, J. V. (1999). Characterization of peptides from Aplysia using microbore liquid chromatography with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry guided purification. *J Chromatogr A* **830**, 105-13.

Fruchter, R. G. and Crestfield, A. M. (1967). The specific alkylation by iodoacetamide of histidine-12 in the active site of ribonuclease. *J Biol Chem* **242**, 5807-12.

Gaskell, S. (1997). Electrospray: Principles and practice (vol 32, pg 677, 1997). *Journal of Mass Spectrometry* **32**, 1378.

Gray, W. R., Olivera, B. M. and Cruz, L. J. (1988). Peptide toxins from venomous Conus snails. *Annu Rev Biochem* **57**, 665-700.

Harvey, D. J., Naven, T. J. P., Kuster, B., Bateman, R. H., Green, M. R. and Critchley, G. (1995). Comparison of fragmentation modes for the structural determination of complex oligosaccharides ionized by matrix- assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **9**, 1556-1561.

Hill, J. M., Alewood, P. F. and Craik, D. J. (1996). Three-dimensional solution structure of mu-conotoxin GIIIB, a specific blocker of skeletal muscle sodium channels. *Biochemistry* **35**, 8824-35.

Hillenkamp, F., Karas, M., Beavis, R. C. and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* **63**, 1193A-1203A.

Hipple, J. A., Fox, R. E. and Condon, E. U. (1945). Detection of Metastable Iond with the Mass Spectrometer. *Physics Reviews* **68**, 54-55.

Hoffmann, E. and Stroobant, V. (2003). Mass Spectrometry: John Wiley & Sons.

Hoving, S., Gerrits, B., Voshol, H., Muller, D., Roberts, R. C. and van Oostrum, J. (2002). Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. *Proteomics* **2**, 127-34.

Huang, L., Baldwin, M. A., Maltby, D. A., Medzihradszky, K. F., Baker, P. R., Allen, N., Rexach, M., Edmondson, R. D., Campbell, J., Juhasz, P. et al. (2002). The identification of protein-protein interactions of the nuclear pore complex of Saccharomyces cerevisiae using high throughput matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry. *Mol Cell Proteomics.* **1**, 434-450.

Jacobsen, R., Jimenez, E. C., Grilley, M., Watkins, M., Hillyard, D., Cruz, L. J. and Olivera, B. M. (1998). The contryphans, a D-tryptophan-containing family of Conus peptides: interconversion between conformers. *J.Pept.Res.* **51**, 173-179.

Jacobsen, R. B., Jimenez, E. C., De la Cruz, R. G., Gray, W. R., Cruz, L. J. and Olivera, B. M. (1999). A novel D-leucine-containing Conus peptide: diverse conformational dynamics in the contryphan family. *J.Pept.Res.* **54**, 93-99.

Jensen, O. N., Podtelejnikov, A. and Mann, M. (1996). Delayed extraction improves specificity in database searches by matrix-assisted

laser desorption/ionization peptide maps. *Rapid Commun Mass Spectrom* **10**, 1371-8.

Jimenez, E. C., Craig, A. G., Watkins, M., Hillyard, D. R., Gray, W. R., Gulyas, J., Rivier, J. E., Cruz, L. J. and Olivera, B. M. (1997). Bromocontryphan: post-translational bromination of tryptophan. *Biochemistry* **36**, 989-994.

Jimenez, E. C., Olivera, B. M., Gray, W. R. and Cruz, L. J. (1996). Contryphan is a D-tryptophan-containing Conus peptide. *J.Biol.Chem.* **271**, 28002-28005.

Johnson, K. L. and Muddiman, D. C. (2004). A method for calculating 16O/18O peptide ion ratios for the relative quantification of proteomes. *J Am Soc Mass Spectrom* **15**, 437-45.

Johnson, R. S. and Taylor, J. A. (2002). Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Molecular Biotechnology* **22**, 301-315.

Jones, R. M. and Bulaj, G. (2000). Conotoxins - new vistas for peptide therapeutics. *Curr Pharm Des* **6**, 1249-85.

Karas, M., Bachmann, D., Bahr, U. and Hillenkamp, F. (1987). Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion. Proc.* **78**, 53-68.

Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-301.

Karas, M. and Kruger, R. (2003). Ion formation in MALDI: the cluster ionization mechanism. *Chem Rev* **103**, 427-40.

Kaufmann, R. (1995). Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *J Biotechnol* **41**, 155-75.

Kaufmann, R., Chaurand, P., Kirsch, D. and Spengler, B. (1996). Post-source decay and delayed extraction in matrix-assisted laser desorption/ionization-reflectron time-of-flight mass spectrometry. Are there trade-offs? *Rapid Commun Mass Spectrom* **10**, 1199-208.

Kaufmann, R., Spengler, B. and Lutzenkirchen, F. (1993). Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun Mass Spectrom* **7**, 902-10.

Keil-Dlouha, V., V, Zylber, N., Imhoff, J., Tong, N. and Keil, B. (1971). Proteolytic activity of pseudotrypsin. *FEBS Lett.* **16**, 291-295.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-43.

Kocher, T., Engstrom, A. and Zubarev, R. A. (2005). Fragmentation of peptides in MALDI in-source decay mediated by hydrogen radicals. *Analytical Chemistry* **77**, 172-177.

Kohn, A. and Hunter, C. (2001). The feeding process in Conus imperialis. *Veliger* **44**, 232-234.

Kostanski, L. K., Keller, D. M. and Hamielec, A. E. (2004). Size-exclusion chromatography-a review of calibration methodologies. *J Biochem Biophys Methods* **58**, 159-86.

Kreil, G. (1997). D-amino acids in animal peptides. *Annu Rev Biochem* **66**, 337-45.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.

Levin, T., Petrides, G., Weiner, J., Saravay, S., Multz, A. S. and Bailine, S. (2002). Intractable delirium associated with ziconotide successfully treated with electroconvulsive therapy. *Psychosomatics* **43**, 63-6.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. and Yates, J. R., III. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat.Biotechnol.* **17**, 676-682.

Lirazan, M. B., Hooper, D., Corpuz, G. P., Ramilo, C. A., Bandyopadhyay, P., Cruz, L. J. and Olivera, B. M. (2000). The spasmodic peptide defines a new conotoxin superfamily. *Biochemistry* **39**, 1583-8.

Ma, B., Zhang, K. Z., Hendrie, C., Liang, C. Z., Li, M., Doherty-Kirby, A. and Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **17**, 2337-2342.

Malmberg, A. B., Gilbert, H., McCabe, R. T. and Basbaum, A. I. (2003). Powerful antinociceptive effects of the cone snail venom-derived subtype-selective NMDA receptor antagonists conantokins G and T. *Pain* **101**, 109-16.

Mann, M. and Wilm, M. (1995). Electrospray mass spectrometry for protein characterization. *Trends Biochem Sci* **20**, 219-24.

Marshall, J., Kelley, W. P., Rubakhin, S. S., Bingham, J. P., Sweedler, J. V. and Gilly, W. F. (2002). Anatomical correlates of venom production in Conus californicus. *Biol.Bull.* **203**, 27-41.

Massilia, G. R., Schinina, M. E., Ascenzi, P. and Polticelli, F. (2001). Contryphan-Vn: A Novel Peptide from the Venom of the Mediterranean Snail Conus ventricosus. *Biochem Biophys Res Commun* **288**, 908-13.

McIntosh, J. M. and Jones, R. M. (2001). Cone venom--from accidental stings to deliberate injection. *Toxicon* **39**, 1447-51.

McIntosh, J. M., Santos, A. D. and Olivera, B. M. (1999). Conus peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu.Rev.Biochem.* **68**, 59-88.

McIntosh, M., Cruz, L. J., Hunkapiller, M. W., Gray, W. R. and Olivera, B. M. (1982). Isolation and structure of a peptide toxin from the marine snail Conus magus. *Arch.Biochem.Biophys.* **218**, 329-334.

Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L. and Burlingame, A. L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **72**, 552-8.

Medzihradszky, K. F., Leffler, H., Baldwin, M. A. and Burlingame, A. L. (2001). Protein identification by in-gel digestion, high-performance liquid chromatography, and mass spectrometry: peptide analysis by complementary ionization techniques. *J Am Soc Mass Spectrom* **12**, 215-21.

Medzihradszky, K. F., Maltby, D. A., Hall, S. C., Settineri, C. A. and Burlingame, A. L. (1994). Characterization of Protein N-Glycosylation by Reversed-Phase Microbore Liquid-Chromatography Electrospray Mass-Spectrometry, Complementary Mobile Phases, and Sequential Exoglycosidase Digestion. *Journal of the American Society for Mass Spectrometry* **5**, 350-358.

Miles, L. A., Dy, C. Y., Nielsen, J., Barnham, K. J., Hinds, M. G., Olivera, B. M., Bulaj, G. and Norton, R. S. (2002). Structure of a novel P-superfamily spasmodic conotoxin reveals an inhibitory cystine knot motif. *J Biol Chem* **277**, 43033-40.

Moczydlowski, E., Olivera, B. M., Gray, W. R. and Strichartz, G. R. (1986). Discrimination of muscle and neuronal Na-channel subtypes by binding competition between [3H]saxitoxin and mu-conotoxins. *Proc.Natl.Acad.Sci.U.S.A* **83**, 5321-5325.

Mok, K. H. and Han, K. H. (1999). NMR solution conformation of an antitoxic analogue of alpha-conotoxin GI: identification of a common nicotinic acetylcholine receptor alpha 1-subunit binding surface for small ligands and alpha-conotoxins. *Biochemistry* **38**, 11895-904.

Morris, H. R., Paxton, T., Panico, M., McDowell, R. and Dell, A. (1997). A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J Protein Chem* **16**, 469-79.

Nakamura, T., Yu, Z., Fainzilber, M. and Burlingame, A. L. (1996). 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* **5**, 524-30.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-21.

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**, 2101-9.

Olivera, B. M., Gray, W. R., Zeikus, R., McIntosh, J. M., Varga, J., Rivier, J., Desantos, V. and Cruz, L. J. (1985). Peptide Neurotoxins from Fish-Hunting Cone Snails. *Science* **230**, 1338-1343.

Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E. E., Woodward, S. R., Hillyard, D. R. and Cruz, L. J. (1990). Diversity of Conus neuropeptides. *Science* **249**, 257-263.

Olivera, B. M., Rivier, J., Scott, J. K., Hillyard, D. R. and Cruz, L. J. (1991a). Conotoxins. *Journal of Biological Chemistry* **266**, 22067-22070.

Olivera, B. M., Rivier, J., Scott, J. K., Hillyard, D. R. and Cruz, L. J. (1991b). Conotoxins. *J Biol Chem* **266**, 22067-70.

Pallaghy, P. K., Melnikova, A. P., Jimenez, E. C., Olivera, B. M. and Norton, R. S. (1999). Solution structure of contryphan-R, a naturally occurring disulfide-bridged octapeptide containing D-tryptophan: comparison with protein loops. *Biochemistry* **38**, 11553-11559.

Patterson, S. D. and Katta, V. (1994). Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry. *Anal Chem* **66**, 3727-32.

Paul, W. and Steinwedel, H. S. (1953). *Naturforschung A* **8a**, 448.

Penn, R. D. and Paice, J. A. (2000). Adverse effects associated with the intrathecal administration of ziconotide. *Pain* **85**, 291-6.

Pereira, C. A., Alonso, G. D., Paveto, M. C., Iribarren, A., Cabanas, M. L., Torres, H. N. and Flawia, M. M. (2000). Trypanosoma cruzi arginine kinase characterization and cloning. A novel energetic pathway in protozoan parasites. *J Biol Chem* **275**, 1495-501.

Porath, J. and Flodin, P. (1959). Gel filtration: a method for desalting and group separation. *Nature* **183**, 1657-9.

Pshenichnyuk, S. A. and Asfandiarov, N. L. (2004). The role of free electrons in matrix-assisted laser desorption/ionization: electron capture by molecules of alpha- cyano-4-hydroxycinnamic acid. *European Journal of Mass Spectrometry* **10**, 477-486.

Reddy, S. R., Houmeida, A., Benyamin, Y. and Roustan, C. (1992). Interaction in vitro of scallop muscle arginine kinase with filamentous actin. *Eur J Biochem* **206**, 251-7.

Rice, R. H., Means, G. E. and Brown, W. D. (1977). Stabilization of bovine trypsin by reductive methylation. *Biochim.Biophys.Acta* **492**, 316-321.

Righetti, P. G. (1989). Isoelectric focussing theory methodology and applications. Amsterdam: Oxford.

Röckel, D., Korn, W. and Kohn, A. (1995). Manual of the Living Conidae. Wiesbaden, Germany: Verlag Christa Hemmen.

Roepstorff, P. (2000). MALDI-TOF mass spectrometry in protein chemistry. *Exs* **88**, 81-97.

Roepstorff, P. and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601.

Sandall, D. W., Satkunanathan, N., Keays, D. A., Polidano, M. A., Liping, X., Pham, V., Down, J. G., Khalil, Z., Livett, B. G. and Gayler, K. R. (2003). A Novel alpha-Conotoxin Identified by Gene Sequencing Is Active in Suppressing the Vascular Response to Selective Stimulation of Sensory Nerves in Vivo. *Biochemistry* **42**, 6904-6911.

Sasaki, T., Feng, Z. P., Scott, R., Grigoriev, N., Syed, N. I., Fainzilber, M. and Sato, K. (1999). Synthesis, bioactivity, and cloning of the L-type calcium channel blocker omega-conotoxin TxVII. *Biochemistry* **38**, 12876-84.

Scheele, G. A. (1975). Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig exocrine pancreatic proteins. *J Biol Chem* **250**, 5375-85.

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. et al. (2001). Two new classes of conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter. *Nat Neurosci* **4**, 902-7.

Shkenderov, S. (1973). A protease inhibitor in bee venom. Identification, partial purification and some properties. *FEBS Lett* **33**, 343-7.

Snyder, L. R., Kirkland, J. J. and Glajch, J. L. (1997). Pratical HPLC Method development. New York: Willey and Sons, Inc.

Snyder, L. R., Stadalius, M. A. and Quarry, M. A. (1983). Gradient Elution in Reversed-Phase Hplc Separation of Macromolecules. *Analytical Chemistry* **55**, 1412.

Stanley, T. B., Stafford, D. W., Olivera, B. M. and Bandyopadhyay, P. K. (1997). Identification of a vitamin K-dependent carboxylase in the venom duct of a Conus snail. *FEBS Lett* **407**, 85-8.

Steen, H. and Mann, M. (2002). Analysis of bromotryptophan and hydroxyproline modifications by high-resolution, high-accuracy precursor ion scanning utilizing fragment ions with mass-deficient mass tags. *Anal.Chem.* **74**, 6230-6236.

Stone, B. L. and Gray, W. R. (1982). Occurrence of hydroxyproline in a toxin from the marine snail Conus geographus. *Arch Biochem Biophys* **216**, 765-7.

Summerfield, S. G. and Gaskell, S. J. (1997). Fragmentation efficiencies of peptide ions following low energy collisional activation. *International Journal of Mass Spectrometry* **165** 509-521.

Sunnerhagen, M., Drakenberg, T., Forsen, S. and Stenflo, J. (1996). Effect of Ca2+ on the structure of vitamin K-dependent coagulation factors. *Haemostasis* **26 Suppl 1**, 45-53.

Sze, E. T., Chan, T. W. and Wang, G. (1998). Formulation of matrix solutions for use in matrix-assisted laser desorption/ionization of biomolecules. *J Am Soc Mass Spectrom* **9**, 166-74.

Szewczuk, A. and Connell, G. E. (1965). The reaction of iodoacetamide with the active center of gamma-glutamyl transpeptidase. *Biochim Biophys Acta* **105**, 352-67.

Talbo, G. and Mann, M. (1996). Aspects of the sequencing of carbohydrates and oligonucleotides by matrix-assisted laser desorption/ionization post-source decay. *Rapid Commun Mass Spectrom* **10**, 100-3.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y. and Yoshida, T. (1988). *Rapid Commun Mass Spectrom* **2**, 151-153.

Taylor, J. A. and Johnson, R. S. (2001). Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Analytical Chemistry* **73**, 2594-2604.

Terlau, H. and Olivera, B. M. (2004). Conus venoms: a rich source of novel ion channel-targeted peptides. **84**, 41-68.

Wagner, S. D., Milstein, C. and Neuberger, M. S. (1995). Codon bias targets mutation. *Nature* **376**, 732.

Wessel, D. and Flugge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal.Biochem.* **138**, 141-143.

Westermeier, R. (1997). Electrophoresis in Practise: A Guide to Methods and Applications of DNA and Protein Separations.: VCH.

Wien, W. (1902). Untersuchungen ü ber die elektrische Entladungen in verdü nnten Gasen. *Annales des Physique* **8**, 244-266.

Woodward, S. R., Cruz, L. J., Olivera, B. M. and Hillyard, D. R. (1990). Constant and Hypervariable Regions in Conotoxin Propeptides. *Embo Journal* **9**, 1015-1020.

Wysocki, V. H., Tsaprailis, G., Smith, L. L. and Breci, L. A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **35**, 1399-406.

Yoon, J. H., Yeo, S. H. and Oh, T. K. (2004). Hongiella marincola sp. nov., isolated from sea water of the East Sea in Korea. *Int.J.Syst.Evol.Microbiol.* **54**, 1845-1848.

Zenobi, R. and Knochenmuss, R. (1998). Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews* **17**, 337-366.

Zhang, Z. Q. (2004). Prediction of low-energy collision-induced dissociation spectra of peptides. *Analytical Chemistry* **76**, 3908-3922.

Zhou, X. D., Jin, Y., Lu, Q. M., Li, D. S., Zhu, S. W., Wang, W. Y. and Xiong, Y. L. (2004). Purification, characterization and primary structure of a chymotrypsin inhibitor from Naja atra venom. *Comp Biochem Physiol B Biochem Mol Biol* **137**, 219-24.