MOLECULAR CHARACTERIZATION OF *Carukia barnesi* and *Malo kingi* . Cnidaria; Cubozoa; Carybdeidae

Thesis submitted by

Griselda Ávila-Soria BSc (Hons) UNAM, MSc UMH

Novembre 2009

for the degree of **Doctor of Philosophy**

in the School of Pharmacy and Molecular Sciences

James Cook University

STATEMENT ON SOURCES

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

…….. ……

(Signature) (Date)

STATEMENT OF ACCESS

I, the undersigned, the author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the copyright act.

I do not wish to place any further restriction on access to this work.

5th of November 2009

Griselda Ávila- Soria Date

DECLARATION ON ETHICS

This research project was carried in compliance with the National Health and Medical Research Council (NHMRC) "National Statement on Ethical Conduct in Research Involving Humans" (1999). **James Cook University Ethics approval number**: **H2022**, granted on 16th of March, 2005.

Field trip collections were done under the **Great Barrier Reef Reserve Park permit number: GO7/24733.1.** Granted to Ávila-Soria, G and Beltran-Ramirez,V.H.

---------- 5th of November 2009

Griselda Ávila-Soria Date

STATEMENT ON THE CONTRIBUTIONS

Financial

The following organizations were involved in this research project:

School of Pharmacy and Molecular Sciences at JCU for a Graduate Research Scheme grant.

Australian Lions Foundation "grant for medical & scientific research into marine species dangerous to humans". Awarded twice.

Queensland Government through CRC Reef (Irukandji Task Force); awarded twice to Professor James Burnell and Griselda Ávila-Soria.

Mexican Council of Science and Technology. Scholarship 196605.

Contribution of others

The following people participated in field trips and were of great assistance or shared expertise with me and Hugo: Surf life Saving Australia, Lisa Gershwin, Jamie Seymour, Ken Moss, Grant Smallwood, all the volunteers during the field trips particularly Ursula Tems. *M. kingi* samples were kindly supplied by Russell Hore. Human sera were provided by anonymous donations in accordance with JCU Ethics Approval Number H2022. Species-specific antibodies raised in mice were carried out under JCU Ethics Approval Number A936 in the laboratory of associate Prof. Heinrich Körner. *C. barnesi* antibodies raised in rabbits and internal amino acid sequences were supplied by K. Winkel (Australian Venom Research Unit, Melbourne University), and rabbit polyclonal antibodies specific for *C. fleckeri* 43/45 KDa toxins (CfTX-1 and 2) were provided by Diane Brinkman (Australian Institute of Marine sciences)

Griselda Ávila-Soria 5th of November

ACKNOWLEDGMENTS

My gratitude to my dear friends Ursula Tems and Diane Brinkman for their support, encouragement, professional advice and to deal so well with my savage and contradictory temper.

I must express my most humble and honest appreciation to Papá James Burnell for being my mentor and to give me the opportunity to continue working with Cnidarians.

To Victor Hugo Beltrán Ramirez my parents and siblings.

THESIS ABSTRACT

This thesis reports the molecular characterization of *C. barnesi* Southcott, 1967 and *Malo kingi* Gershwin, 2007; two Australian box jellyfish responsible to trigger, a complex and variable illness in humans, termed -Irukandji syndrome-. This is characterised by a 30 min delay before severe systemic symptoms occur including severe pain, catecholamine effects and in some cases cardiopulmonary decompensation. Between 100 to 200 cases of Irukandji syndrome are reported each year and in 2002, Irukandji jellyfish caused the death of two people in North Queensland, Australia. *M. kingi* is presumed to be responsible for one of these fatalities. Currently there is no cure or antivenom for Irukandji box jellyfish stings. and studies have been hampered by the Irukandji small size and seasonality. In addition, sparse Cubozoa taxonomy has contributed to a general lack of knowledge of these medusae.

Samples of Irukandji box jellyfish were collected during the 2003 and 2005 summers at North Queensland beaches and on the Great Barrier Reef. The mRNA of one specimen of *C. barnesi* was *in vitro*-amplified and used for the construction of an expression cDNA library, which was then amplified once. The mRNA of two adult *M. kingi* specimens was combined to generate a non-normalized cDNA library.

The cDNA libraries constructed were immunologically screened with species-specific antibodies, human sera and *Chironex fleckeri* antivenom. The antibodies were antigenic against native cubozoan venom proteins, but failed to specifically bind venom proteins expressed in bacteria. However, the antibodies were reactive against several proteins exhibiting structural, catalytic or chaperone activity. Although these results were unexpected, they forced us to reconsider our early perceptions of some of the toxic protein properties.

The *M. kingi* cDNA library proved to be a good quality molecular tool that allowed the establishment of a well-characterized and non-redundant EST resource from which were identified novel transcripts, several serine and zinc proteinases and their inhibitors, a phatogenic like gene, allergens, two neurotoxin-like genes (CbX and Mk-332), two cytolysins: MkTX-A and B homologous to those previously reported. In addition, several of the encoded proteins were expressed in a bacterial system and further characterized. Major findings not only included the identification of highly expressed defence genes but also, localization of several of these using *in situ* hybridization techniques indicated their putative involvement in the box jellyfish defence system.

During the analysis of the Cubozoan sequence data, a significant number of genes displaying high similarities with genes from plants, prokaryotes, viruses and even unicellular eukaryotes were also identified. Surprisingly, some genes were phylogenetically more closely related to higher animals such as primates than to invertebrate animals similar to jellyfish such as ctenophores or sponges. It was interesting to find that Cubozoans, simple, brainless, and venomous animals, display such an elevated level of gene sophistication.

Jellyfish research has been previously hampered due to limited availability of biological specimens, and this was overcome during this work. The molecular information presented in this thesis represents a basis for future investigations.

LIST OF MOST USED ABBREVIATIONS

Amp	Ampicillin
Abs	Absorbance
bp	Base pairs
BSA	Bovine serum albumin
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
Da	Dalton
C	Carbenicillin
cDNA	Complementary deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
EtBr	Ethidium bromide
g	Grams
HRP	Horse radish peroxidase
IPTG	Isopropyl β-D-thiogalactopyranoside
Kb	Kilobases
kDa	Kilodaltons
kg	Kilograms
Km	Kanamycin
LB	Luria Bertani media
Min	Minute
M	Molar
mM	Millimolar
μM	Micromolar
nM	Nanomolar
pM	Picomolar
mSec	Millisecond
M.W	Molecular weight
nm	Nanometers
ORF	Open reading frame
OD	Optical Density
PBS	Phosphate buffered saline
rpm	Revolutions per minute
Sec	Seconds
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Pfu	Plaque forming unit
PVDF	Polyvynlidene difluoride
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
^{o}C	Degrees celsius
UTR	Untranslated region

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

BACKGROUND

The focus of this work was the molecular characterization of proteins from Irukandji box jellyfish. These marine creatures belong to the early diverging metazoan phylum Cnidaria; a diploblastic lineage, that stands as the sister group of triploblastic animals and are considered the first organisms to evolve with a nervous system (Anderson *et al.* 1993; Lindgens *et al.* 2004). Moreover, it is one of the most venomous groups in the animal kingdom.

The phylum comprises benthic and pelagic aquatic animals including four living classes (see Figure 1.1): anthozoans at the basal position, represented by hard corals, soft corals, sea fans and anemones; hydrozoans such as the legendary fluorescent *Aequorea victoria* and the deadly hydroid *Physalia physalis* and scyphozoans, (e.g. *A. aurita*, R. *nomadica, C. capillata*) which are all true jellyfish; and cubozoans, also known as box jellyfish (Daly *et al.* 2007; Nevalainen *et al.* 2004).

Cnidarians display radial symmetry, composed of an ectoderm and mesoderm, both of which constitute the body wall, and a middle collagen layer called the mesoglea (Figure 1.2). Ectodermal and endodermal cell lineages are distinct. Epithelial cells play a crucial role in several physiological functions; protective functions and osmoregulation when ectodermal and digestion when endodermal (Ruppert *et al.* 2004). Muscular contraction occurs in both ectoderm and mesoderm. According to the differentiation state, interstitial cells will differentiate into one of the following: gametes, secretory cells such as glands, mucous cells and either of two types of neuronal cells, ganglia or cnidocytes, which are highly specialized combinatory sensory-effector cells. These unique cells are a defining feature of Cnidarians and play a central role in prey capture and defence (Anderluh *et al.* 2000b; Ruppert *et al.* 2004) .

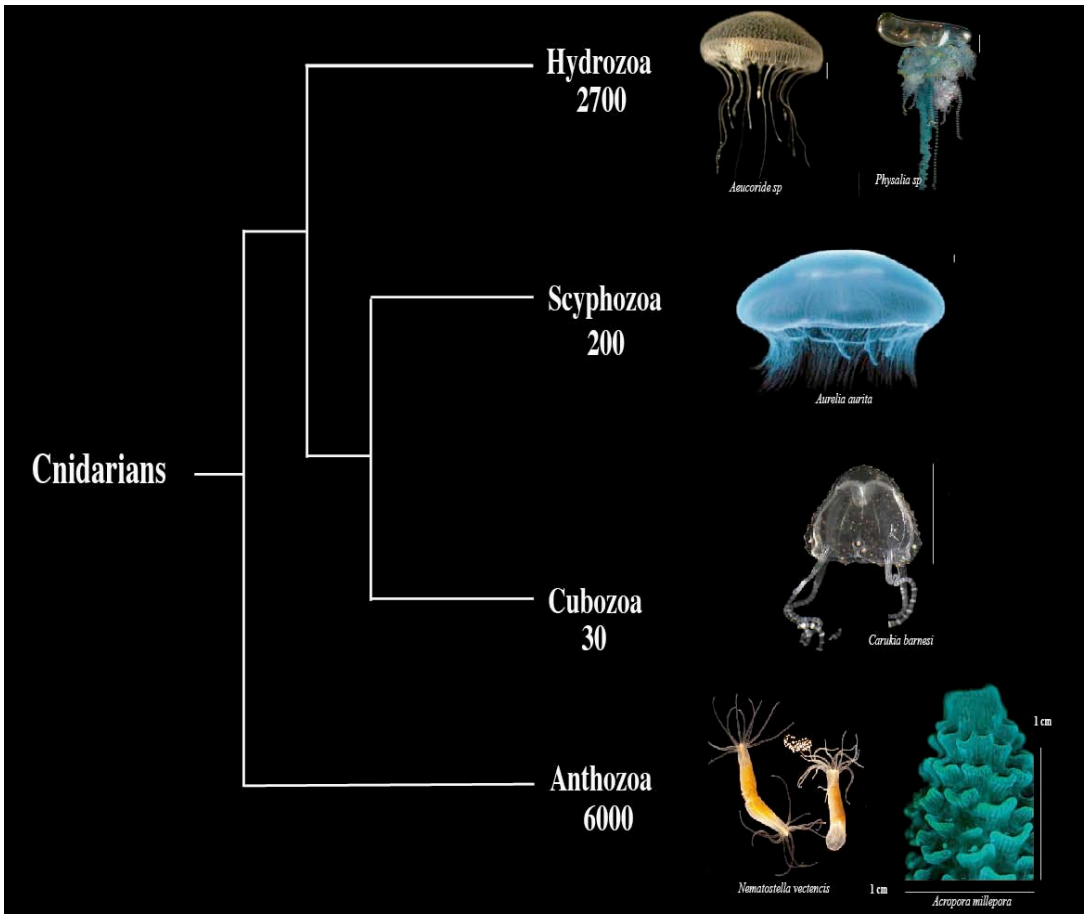


Figure 1. 1 Cnidarian class-level relationships

The tree is redrawn and modified after Bridge *et al.* (1995). The branch belongs to the staghorn coral and the polyp to *N. vectencis*, both of which are anthozoans. The zig-zag and blue bottle jellyfish are hydrozoans, the moon jelly is a true jellyfish and the box jellyfish is an Irukandji. Photographs taken by (*A. millepora*) V.H. Beltrán, (Hydrozoans) L. Gershwin, (*A. aurita*) G. Matsumoto, (*C. barnesi*) G. Ávila-Soria and (*N. vectencis*) is courtesy of J.R. Finnerty.

The cnidocytes house a unique explosive organelle called the cnidocyst, produced through a complex secretory pathway (Anderluh *et al.* 2000b). Three types of cnidocysts are known: nematocytes or "stinging cells" aimed for injection of venom, spirocytes for construction of felt-like tubes and ptychocytes for adhesion to substrates and prey (Mariscal 1974). Only nematocysts are of medical importance and these are found across all four cnidarian classes (Daly *et al.* 2007), whereas ptychocysts, and spirocysts are exclusive to anthozoans (Williamson 1992) and have functions other than defence or predation.

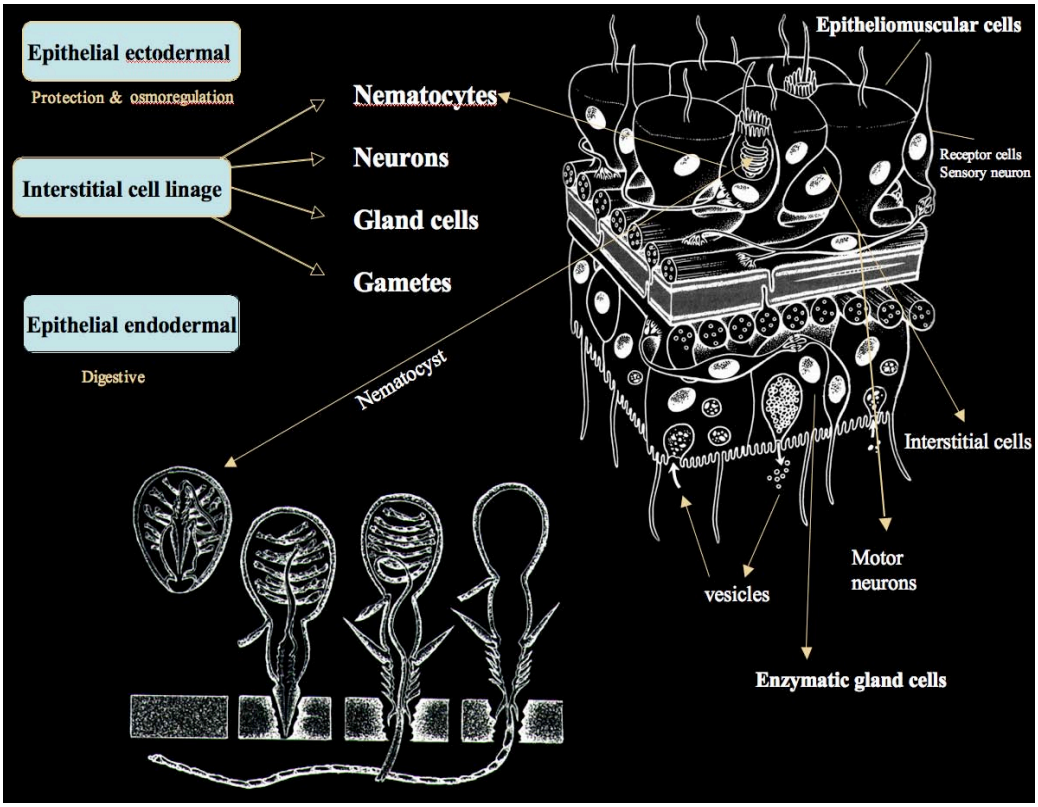


Figure 1. 2 Cnidaria body plan and nematocyst dynamics

A) Tissue layer organization and interstitial cell biogenesis. Adapted from (Ruppert *et al.* 2004). B) Nematocyst discharge and prey tegument penetration. Adapted from (Tardent & Holstein 1982).

At least 24 cnidocyst types exist in cnidarians but mastigophores, isorhizas and euryteles are the main types of nematocysts found in the cubozoans [see Gershwin (2005) for detailed information on Class Cubozoa].

Considering their small size, nematocysts are highly structured organelles. They are composed of a thick walled collagen-sealed capsule (Engel *et al.* 2001) containing an inverted tubule lying coiled and folded and connected with a stylet (Lotan *et al.* 1996). The capsule interior is also surrounded by fluid that contains amino acids, peptides and a high concentration of ions. Many of the ions (K^+, Mg^{2+} and especially Ca^{2+}), are complexed with macromolecules, forming large aggregates that create little osmotic pressure, but when sensory-like hairs on the outer surface of the nematocyte are mechanically or chemically stimulated, these ions dissociate creating a high intra-capsular osmotic concentration. This causes water from the cytoplasm to rush into the capsule, creating a high hydrostatic pressure (Holstein *et al.* 1994; Tardent 1995).

The tubule can eject completely within 3 milliseconds (Ruppert *et al.* 2004), with an acceleration of $1x10^6 g$ and a force of skin penetration of around 25 kPa (Lotan *et al.* 1996; Tardent 1995). The stylet can penetrate into human skin up to 0.9 mm, deep enough to deposit a blend of toxins into the microvasculature of the dermal tissue to be absorbed into the systemic circulation. This impressive mechanism is used by cnidarians to capture their prey as well as a system of defence against predators (Mariscal, 1974; (Hellstern *et al.* 2006; Watson & Mirethibodeaux 1994).

1.1 Biological activity displayed in Cnidarians

The complex blend of proteins and peptides located in the nematocysts of all cnidarians exhibit biological activity (Bloom *et al.* 1998; Burnett *et al.* 1992; Chung *et al.* 2001) in several test systems. Some of the toxic proteins display their poisonous effect by targeting and modifying the properties of voltage-gated ion channels (Messerli & Greenberg 2006) thus producing cytotoxic (Endean *et al.* 1993; Neeman *et al.* 1979), myotoxic and/or cardiotoxic effects (Walker *et al.* 1977a). Phospholipase A2 (Lotan *et al.* 1996; Nevalainen *et al.* 2004; Radwan *et al.* 2005; Talvinen & Nevalainen 2002; Walker *et al.* 1977a) haemolytic (Gusmani *et al.* 1997; Nagai 2003; Rottini *et al.* 1995), haemotoxic (Azila *et al.* 1991), proteolytic (Hessinger & Hessinger 1981; Calton & Burnett, 1982), antitumoral (Moussoukhoye & Abdoulaye 2004) and antimicrobial activities of cnidocyst contents have also been reported (Morales-Landa *et al.* 2007; Ovchinnikova *et al.* 2006).

In addition, other biologically active compounds in cnidarians may include neurotransmitters such as vasoactive serotonin-like (5-hydroxytryptamine, or 5-HT) (Bouchard *et al.* 2004), catecholamines, histamine-like substances and histamine liberators and prostaglandins (Gerhart 1991).

1.2 Human envenomation by cnidarians

Of the ~10,000 species of cnidarians, at least 100 of them cause damage to humans by direct contact with their nematocysts or by ingestion. Cnidarian venom components

vary from one species to another in potency and toxic effects produced in humans (Burnett *et al.* 1986). The damage inflicted by cnidarians can be caused by immune reactions that are triggered when toxins enter organisms, or by the poisonous effect generated by the same toxins. The cnidarian venoms are toxic mainly to the neuronal, cardiovascular, respiratory and renal systems, which can cause the appearance of systemic symptoms accompanied by cutaneous injuries, displaying gangrene, contractions, neuritis, hyper-pigmentation, ulcerations in the cornea and effects similar to post-intoxication granulomas or herpes simplex (Burnett 1991; Burnett *et al.* 1996), paralysis and death.

Within cnidarians, the subphylum Meduzoa is the most medically important as several species cause public health problems on a global scale; some species are among the most venomous organisms known (Halstead, 1978). These include mainly members of Classes Scyphozoa and Cubozoa.

Scyphozoans, such as *Catostylus mosaicus* (Wiltshire *et al.* 2000) *Pelagia noctiluca* (Malej & Malej 1987; Mariottini *et al.* 2002), *Chrysaora quinquecirrha* (Bloom *et al.* 2001; Neeman *et al.* 1979) *Chrysaora achlyos* (Radwan *et al.* 2000), *Cassiopea xachamana* (Radwan *et al.* 2005; Torres *et al.* 2001), *Rhopilonema nomadica* (Yu *et al.* 2007), *Aurelia aurita* (Ávila-Soria 2001; Nagai *et al.* 2002; Noguchi *et al.* 2005; Radwan *et al.* 2001; Segura-Puertas 2002) have been reported to inflict damage on humans.

Cubozoans are generally much more toxic than scyphozoans. For example, in Australia alone, more than 70 deaths have been reported due to *C. fleckeri* stings and *Chiropsalmus quadrigatus* has caused fatalities in Japanese seawaters (Nagai *et al.* 2002; Noguchi *et al.* 2005). Some other cubozoan species that have caused severe injuries to humans include *Carybdea marsupialis* (Estrada-Muñoz 2001; Sanchez-Rodriguez *et al.* 2006), *Carybdea alata* (Chung *et al.* 2001), *Carybdea rastoni* (Nagai 2003) and finally the Irukandjis such as *C. barnesi* (Barnes 1964; Fenner 2006), *Carukia shinju*, *Malo maxima* (Gershwin 2005) and *M. kingi* (Gershwin 2007a) which are likely to cause death.

1.3 Class Cubozoa

Phylogenetic relationships among the medusozoan are generally poorly resolved. Until 1973, Cubozoa were grouped together within the Class Scyphozoa (Arneson & Cutress 1976). However specific anatomical structures and reproductive strategies and life cycle differences between the classes exist (Werner 1975), indicating that they are taxonomically different and supporting Cubozoa at the class level. Within Class Cubozoa there are two orders; Chirodropida and Carybdeida.

Cubomedusae have the most complex neurosystem and display intricate feeding and aggressive hunting behaviour. They are agile swimmers and remarkably they have evolved well-developed image-forming eyes. In spite of their enhanced competitive

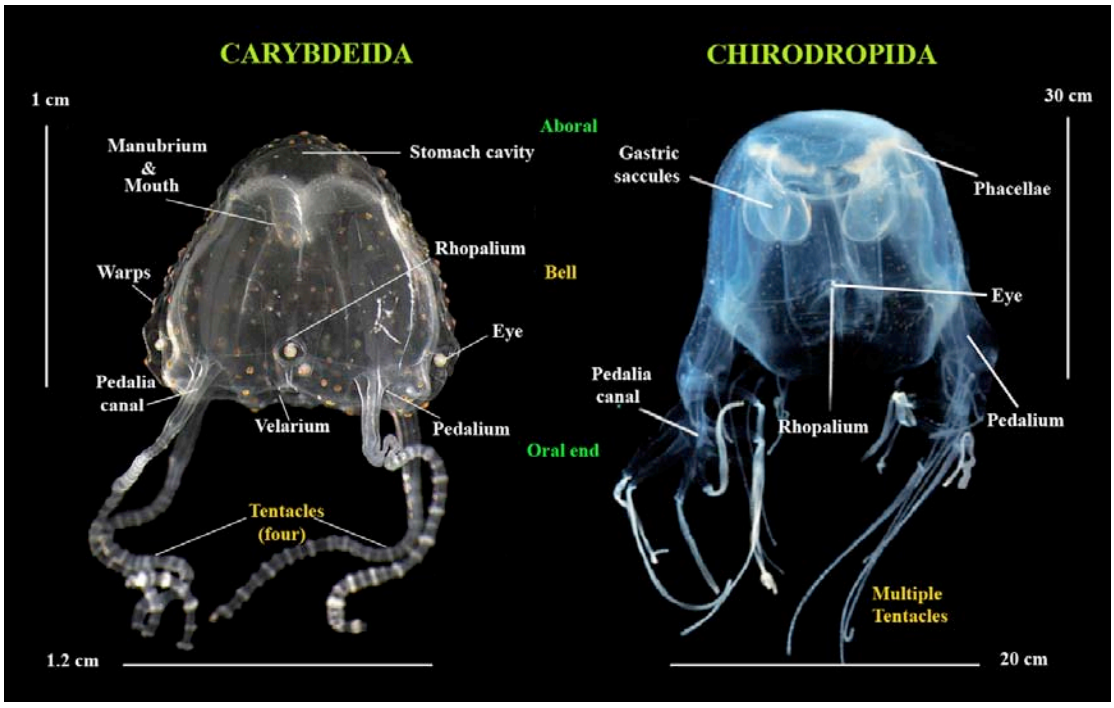


Figure 1. 3 Carybdeida and Chirodropida morphological differences

Carybdeids have characteristic blade-like pedalia, each containing one tentacle, and no gastric saccules. Chirodropids have forked pedalia, each "finger" containing a tentacle and gastric saccules. Note the differences in size and tentacle number. Carybdeids are considerably smaller and four-tentacled with the exception of *Tripedalia cystophora* (with three tentacles at each corner). In contrast, *Chironex fleckeri* can be 30 cm wide and possess up to 15 tentacles per corner (Photograph by L. Gershwin).

abilities, cubozoans are the most species-poor class, containing approximately 36 species (see Figure. 1.1 for comparison). Cubozoans or box jellyfish are distinguished from other medusae by their body pattern, which is a hollow umbrella-like body that is square in shape. The manubrium and mouth are located inside the umbrella and the velarium is situated along the edge of the umbrella. Tentacles are attached to pedalia, which are muscular extensions of the umbrella. Morphological differences between the two box jellyfish orders are more easily observed in Figure 1.3. Box jellyfish are found in tropical regions around the world. In particular in Australia, they are commonly found between November and May in northern coastal waters.

1.4 Irukandji syndrome-causing jellyfish

Irukandji box jellyfish are named after an aboriginal tribe which inhabited the coastal region between the Mowbray River to the north of Trinity Inlet and around Cairns to the south (Barnes 1964) - see the map in Figure 1.4. The first Irukandji jellyfish species was discovered in Australian north tropical waters more than 56 years ago (Flecker 1952). The syndrome that follows envenomation by Irukandjis is a collection of symptoms, characterised by slight local pain, sometimes the skin may develop redness but in all cases there is a delay of approximately 30 minutes before the onset of severe systemic symptoms. These include pain, catecholamine effects and in some cases cardiopulmonary decompensation. Affected individuals may have a combination of these symptoms, but consistently suffer pain (Fenner & Carney, 1999). In more severe cases intracerebral haemorrhage has been observed after severe hypertension, causing death. Two deaths in 2002 from Irukandji syndrome have been reported: one in the Whitsunday Islands, and the second in the tropical Great Barrier Reef region (Fenner & Hadok 2002), see figure 1.4.

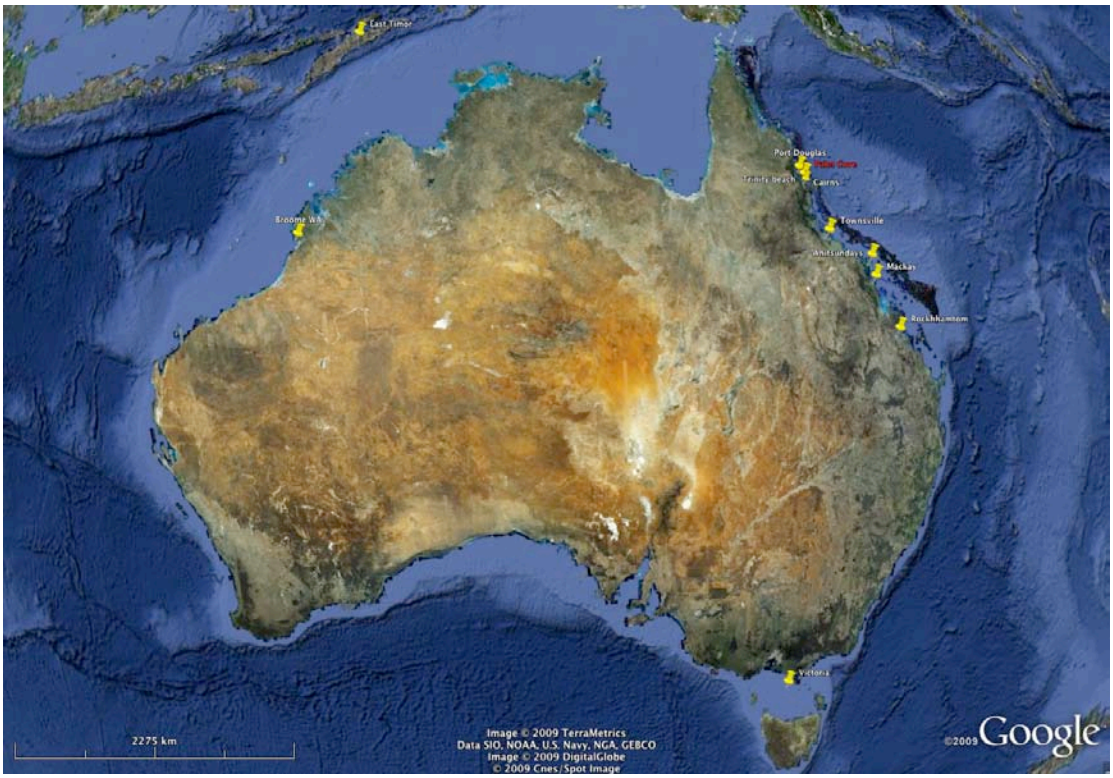


Figure 1. 4 Map of Australia indicating reports of Irukandji syndrome

Note that *C. barnesi* and *M. kingi* species had been reported to inhabit waters only in North Queensland.

Irukandji envenomations are not limited to North Queensland (Kinsey *et al.* 1988), nor restricted to Australia. Cases have been reported in Broome (Marsh *et al.* 1986a) and in Australia as far south as Victoria (Cheng *et al.* 1999) and throughout tropical and temperate oceanic regions of the world such as Bon Air in Florida, (Grady & Burnett 2003), French West Indies in the Caribbean (Pommier *et al.* 2005), Thailand (De Pender *et al.* 2006), Hawaii, East Timor and Papua New Guinea (Little *et al.* 2006)

Until 2002, the only species identified as causing the syndrome was *C. barnesi* (Fenner & Hadok 2002), however recent research suggests that this syndrome is caused by at least six small carybdeid jellyfish similar to *C. barnesi*. Three of these have recently been described and include *C. shinju*, *Malo maxima* (Gershwin, 2005) and *M. kingi* (Gershwin 2007a). *Alatina mordens* (Gershwin, 2005) has also been

linked with cases of Irukandji syndrome and Morbakka (Tamoya) has been demonstrated to cause a mild symptomology.

According to some authors (Huynh *et al.* 2003), the nematocysts obtained from one of the 2002 victims were not from *C. barnesi*, but identified as a unique nematocyst type implicating *M. kingi* as the most likely cause of the death (Gershwin 2005a; Gershwin 2006b; Gershwin 2007b).

The main objective of this thesis was to characterize, at the molecular level, *C. barnesi* and *M. kingi*. Both species belong to the order Carybdeida, have four tentacles and a characteristic cuboidal shape that narrows slightly towards the apex. Figure 1.5 highlights the general characteristics of *C. barnesi*. For example, the adult medusa is small with a length of 1.5 to 2.5 cm with prominent red mamillations (clumps of nematocysts) over the exumbrella as well in the velarium. *C. barnesi* is distinguished from other Carybdeidae by the absence of gastric phacellae, which are usually present at each corner, and can be 5 to 7 cm long when contracted and almost 1 m long when extended. The nematocysts of the tentacles are mainly homotrichous microbasic rhopaloids, with the capsule an average of 25 to 26 µm long and 15 to18 µm wide. Nematocyst bands in the genus *Carukia* possess a distinctive neckerchief form, so named by Southcott (1967). The mamillations or warp cup at the umbrella contain approximately 20 to 25 nematocysts that are classified as haploneme anhisorhizas and possibly homotrichous. *C. barnesi* is endemic to Australia (typically from Port Douglas in North Queensland to the Whitsundays near Mackay) in coastal and oceanic waters (Figure 1.4).

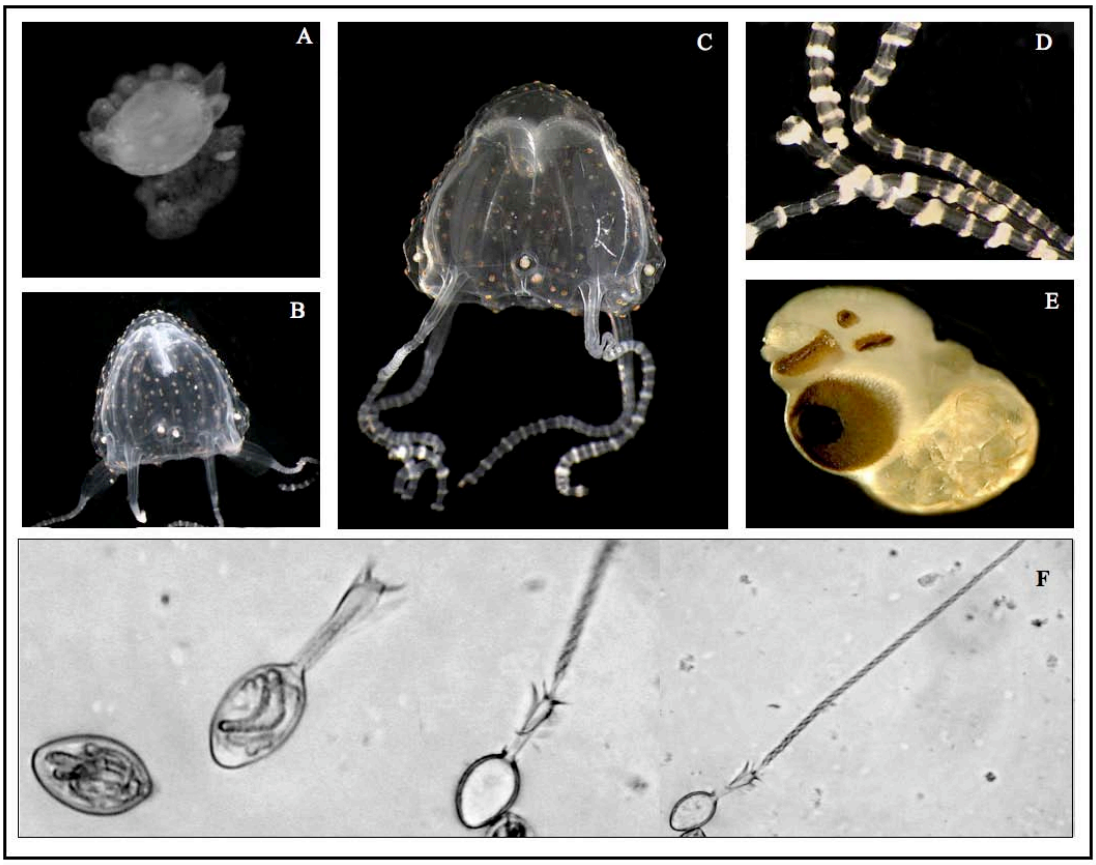


Figure 1. 5 Microphotographs of *C. barnesi* and nematocysts.

A. Cultured polyp; B. Juvenile stage; C. Adult specimen; D. Elongated tentacles; E. Ropalium, statoliths and eyes; F. Tentacular nematocyst discharge process. (Photographs A-D and F by G. Ávila-Soria. Photograph E by L. Gershwin).

In contrast, *M. kingi* is a much more robust jellyfish compared to *C .barnesi*. *M. kingi* adults can grow in height from 2.5 to 2.8 cm and 2.3 to 2.8 cm wide (Figure 1.6). In living specimens, the umbrella is transparent and colourless; the tentacles sometimes pink with pale purple exumbrellar nematocyst warts. *M. kingi* differs from all other cubozoans by the presence of halo-like rings of tissue encircling the tentacles, with nematocysts inserted end-on around the periphery of the rings. The tentacular nematocysts are elongate club-shaped type 4 microbasic p-mastigophores that are 30.27 to 36.68 μm long and 13.02 to 16.04 μm wide. The exumbrellar nematocysts are spherical isorhizas, 20.10 to 24.87 μm in diameter (Gershwin 2007b). Little is known about *M. kingi* ecology. Samples have been found in late summer and early winter. To my knowledge only five jellyfish of this species have been found; one at Mackay and four at Port Douglas (refer to Figure 1.4), three of those specimens were allocated to this research.

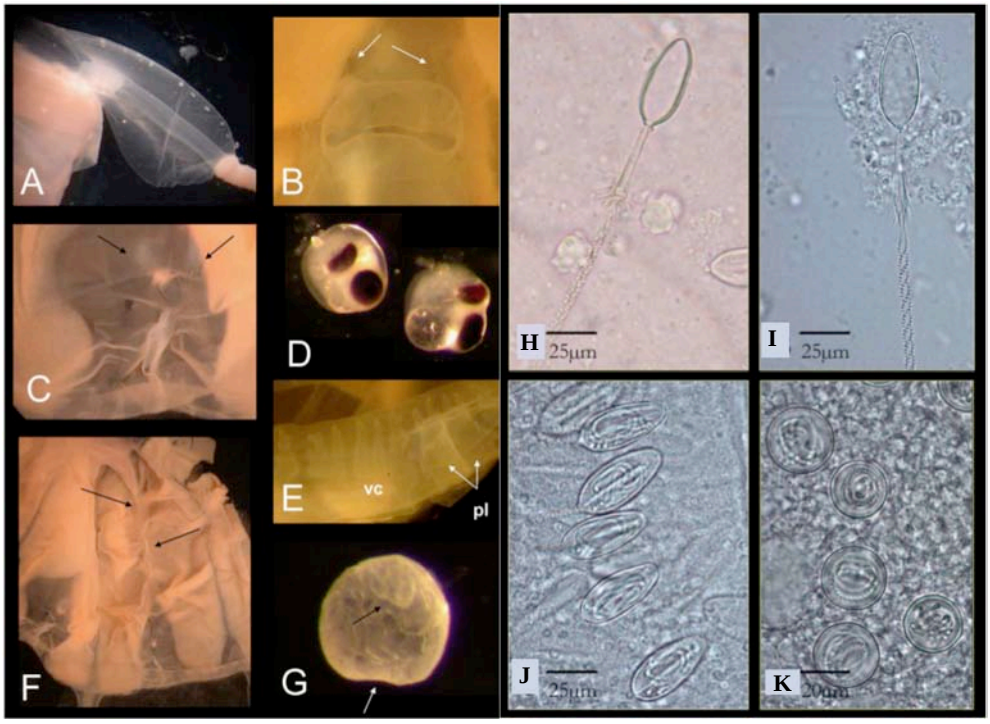


Figure 1. 6 Microphotographs of *Malo kingi* general morphology

A. Pedalium; B, C and D. Rhopalia; E. Velarial canals; F. Mesenteries; G. Statolith; Tentacular and umbrellar nematocyst; H and I. Type 4 microbasic p-mastigophores, discharged, tentacles; J. Club-shaped microbasic p-mastigophores charged, tentacle; K. Spherical isorhizas discharged, umbrella. Images from (Gershwin 2007b).

1.5 Recent cubozoan research

In the last 10 years important progress has been achieved in the molecular, biochemical and toxicological characterization of venom proteins from cubozoans. A novel cytolytic protein family comprising six polypeptides, identified from four box jellyfish (two from each cubozoan order): *Chironex fleckeri*, *Chiropsalmus quadrigatus*, *Carybdea rastoni* and *Carybdea alata*, have been identified. The box jellyfish venom proteins have haemolytic properties, are lethal to crabs, crayfish and mice and produce dermonecrotic effects in the skin of mice (Brinkman & Burnell 2007; Nagai 2003; Nagai *et al.* 2000a).

All of the box jellyfish cytolytic proteins are found to be major constituents of nematocyst venom. The corresponding genes have been cloned and comparisons of

the deduced primary and predicted secondary structure of these proteins revealed a relatively low evolutionary rate operating in compact transmembrane spanning regions. This indicates a high functional importance of these conserved blocks that putatively may be involved in the lytic action mechanism. In addition, a large variation from site to site among the homologous box jellyfish cytolysins is seen, indicating that the functional importance of the semi conserved sites may change over time. This also suggests a possible functional divergence among these related proteins. In addition, Brinkman and Burnell (2009) reported two other major proteins (39 and 41 kDa) and a minor cytolytic protein identified and partially purified from *C. fleckeri*. Although important contributions have been achieved, none of the reported proteins come from cubomedusae eliciting Irukandji syndrome in humans.

In the last six years there has been great interest in Irukandji syndrome-causing jellyfish, since the death of a North American tourist, which occurred on the Great Barrier Reef in 2002. Programs have been implemented such as the Irukandji Task Force, an initiative of the Queensland Government and the Lion Foundation support for "Scientific research into marine species dangerous to humans". These initiatives have greatly shaped the landscape in box jellyfish research, resulting in increases in knowledge of cubozoan ecology, biology, toxinology and also in fundamental molecular biology. This research was initiated to shed some light on the defence mechanism genes in cubozoans, responsible for the development of Irukandji syndrome following envenomation.

RATIONALE OF THE STUDY.

Cnidarians represent an important phylum in nature due to their evolutionary position with respect to bilaterian animals. In addition, they are also medically important due to the threat to human health that many members represent. The choice of *C. barnesi* as the main study target was due to its predominance as a public health hazard and the absence of clear, specific information on the biochemistry and pharmacological properties of its venom. The project was broadened to include study of *M. kingi*, which was chosen following its implication as the cause of a fatality in 2002. In contrast, no fatality has been reported due to *C. barnesi* and on this basis *M. kingi* appears to be more venomous than *C. barnesi*, so a comparison of the two species at a molecular level was an appealing prospect.

CHAPTER 2

Immunoscreening of cDNA expression libraries from *Carukia barnesi* and *Malo kingi* and partial characterization of Irukandji proteins

2.1 ABSTRACT

Carukia barnesi (Southcott 1967) and *M. kingi* (Gershwin 2007) are small, rare and dangerous box jellyfish found in the coastal waters of North Queensland, Australia, during the summer months. These two jellyfish vary in the potency of the toxic effects produced in humans and both are responsible for the so-called "Irukandji syndrome" (refer to Chapter 1). In an attempt to identify and characterize venom proteins present in Irukandji box jellyfish, a study using molecular biological tools was undertaken. The study involved the use of antibody probes to screen *C. barnesi* and *M. kingi* cDNA expression libraries, constructed in a Lambda Zap II vector. A number of antibodies were employed to independently screen the libraries: CSL box jellyfish antivenom (CSL AV) produced in hyperimmunised sheep against *C. fleckeri* venom (CSL Ltd., Australia), sensitized human sera and species-specific antibodies (raised in rabbit and mice). Following consecutive rounds of immunoscreening, positive clones were identified, isolated, sequenced and annotated by comparing sequence data to GenBank.

Results of this study indicate that complex antibody probes are not suitable to detect Irukandji toxic components expressed in cDNA expression libraries. The data show unwanted background and no detection of genes encoding proteins similar to venom proteins in other metazoans. However clones encoding proteins such as Cyclophilin A were found, which are homologous to those that act as allergens in other organisms. Furthermore, these experiments provided valuable transcriptional information for two cubozoans from which no molecular data was previously available.

Key words. Irukandji box jellyfish; CSL box jellyfish antivenom (CSL AV); antigenicity; Human sera; nematocyst extracts; cDNA expression libraries.

2.2 BACKGROUND

The Irukandji syndrome is named after an aboriginal tribe that inhabited the Palm Cove region where this envenomation was commonly reported (Barnes 1964; Flecker 1952). The illness was first described in Cairns by Hugo Flecker in 1952 and the agent causing the illness (*C. barnesi*) discovered in 1964 by Jack Barnes (Barnes 1964; Kinsey *et al.* 1988). Each year between 100 and 200 cases of Irukandji syndrome are reported, and two fatalities occurred in 2002 (Fenner & Hadok 2002). Envenomation by the recently described species *M. kingi* is considered to be the cause of one of the fatalities (Gershwin 2007b).

A key attribute of cubozoans is the presence of cnidocytes or cnidae cells controlled by the nervous system. Each of these "single-use" cells accomplishes only one purpose: to produce the nematocyst, a sophisticated mechanical sensor capsule-like attack device that has a wide variety of sizes, shapes and barbs. The lumen of the capsule contains a complex arsenal of peptides and polypeptides directed to trap, immobilize and ultimately kill prey. Nematocysts are located in the tentacle arranged in batteries. Some cubozoans, specifically Irukandji species, also contain random nematocyst clusters on the umbrella. Cubozoan venom is likely to contain toxic as well as non-toxic components and the amount and types of these may indicate differences among species within the class.

Valuable knowledge in terms of isolation and structure elucidation has been gained through classical biochemistry studies. Research of toxic peptides and polypeptides from Cnidarians has been focused on dominant hydrozoan species as well as anthozoans due to sessile habits, accessibility, and ability to be kept in laboratory conditions. Still there are abundant studies on scyphozoans and to a lesser extent on cubozoans. There have been several difficulties encountered in these studies such as high molecular weights of toxic proteins, high protease activity, aggregation, disaggregation, surface adhesion, and thermolability (Abe *et al.* 1999; Bloom *et al.* 1998; Chung *et al.* 2001; Nagai 2003; Rottini *et al.* 1995; Wiltshire *et al.* 2000). In addition, the inpurity of venom preparations (Ramasamy *et al.* 2005b), e.g., symbiont, mucus and tissue contamination and also inaccurate species identification (Gershwin

2006b) has led to misinterpretations of results and handicapping of research progress (Tibballs, 1998, Currie 2003).

In Australia, the best characterized Cubozoan has been the chirodropid *C. fleckeri* due to its abundance, size and numerous tentacles and its tentacles contain millions of nematocysts and potent, fast-acting toxins. This box jellyfish is responsible for most of the marine envenomations in northern Australia, and although in the majority of the cases, *C. fleckeri* envenomation is not life threatening (Bailey, 2003), more than 70 people have died due to *C. fleckeri* stings in the last decade (Ramasamy *et al.* 2004) The two most recent victims were a seven year old aboriginal girl in far north Queensland, January 2006 (ABC News 2006), and a 6-year old boy from Tiwi Islands in the northern territory (ABC News 2007).

On the other hand, there are fewer studies of venom compounds affecting humans from seasonal, rare, small, free–living and four-tentacled Caribdeidae box jellyfish, such as members of the Irukandji-causing group. This is mainly due to the hazards of collection, their small size and camouflage (no evident coloration), making specimens difficult to find. Like all the members of Class Carybdeidae, box jellyfish have high content of water (~approximately 95.5 %). As a result, very little information has been gathered on the biological activity and action mechanism of toxic proteins from Irukandji. At the time of writing, only two research papers devoted to the Irukandji venom action mechanisms were available: Winkel and collegues (2005) reported the cardiovascular effects in human, rat and guinea-pig tissues (*in vitro)* and in anaesthetized piglets (*in vivo)* of whole *C. barnesi* venom. It was found that Irukandji venom causes tachycardia and systemic pulmonary hypertension, triggered by a massive release of neurotransmitters – catecholamines. This result is in part reinforced by a second study regarding the *in vivo* effects of nematocyst proteins from *C. barnesi* as well as tentacle extracts devoid of nematocysts tested in anaesthetized rats. The result was production of a cardiovascular pressure response that was not dose dependent. The difference between the two extracts was that nematocyst proteins produced cardiovascular collapse whereas the tentacle tissue did not, and the response was slower (Ramasamy *et al.* 2005b). The experimental results of both research groups are consistent with elevated levels of catecholamines circulating in the

bloodstream of Irukandji victims (Fenner *et al.* 1988; Williamson *et al.* 1996; Tibballs *et al.* 2001)**.** Although experimental results explain some of the symptoms observed in Irukandji envenomation, the component targeting an ion-dependent receptor protein of neuroendocrine cells (in the adrenal glands) responsible for the normal balance of catecholamine secretion, has never been found, purified or characterized. If identified, the component could be an important therapeutic target for treatment of degenerative disorders of the nervous system, in particular Parkinson's disease.

2.2.1 Immunological screening of expression cDNA libraries

Expression libraries are an alternative strategy for the rapid identification of cross reactive allergenic and toxic proteins. In addition this system has proven to be useful for overcoming limited material availability and provides a means by which large quantities of fusion proteins can be produced, tested and functionally categorized. Many cDNA sequences encoding toxic proteins have been identified by screening cDNA expression libraries using either specific/unspecific polyclonal or monoclonal antibodies. These include cDNA sequences from plants (Mittermann *et al.* 2005), fungi: *Cladosporium herbarum* and *Aspergillus fumigatus* (Achatz *et al.* 1995; Saxena *et al.* 2003), parasitic invertebrates (Achatz *et al.* 1995; Bachrach *et al.* 1997), honeybees (Whitfield *et al.* 2002) hornets (Fang *et al.* 1988), parasitic wasps (Jones *et al.* 1992) and snakes (Assakura *et al.* 2003; Welton & Burnell 2005).

Several novel and homologous ribosomal proteins that cause allergic reactions and invasive respiratory disorders in humans have been identified. By using the pooled sera of allergic patients with elevated levels of IgG and IgE antibodies, these ribosomal proteins were characterized through screening an *A. fumigatus* cDNA library (Saxena *et al.* 2003). The proteins were similar to 60S ribosomal protein L3 and other ribosomal proteins such as *A. fumigatus* P2, evidenced by ELISA and immunoblotting. Furthermore, recombinant ribosomal protein P2 is able to induce skin reactivity (Mayer *et al.* 1999). In *Brucella melitensis* and *Cladosporium herbarum,* ribosomal proteins L7 and L12, were also reported to be allergens (Achatz *et al.* 1995; Bachrach *et al.* 1997). Furthermore, in plants a number of cross reactive allergens, responsible for pollen and food allergies, are composed of cytoskeletal

proteins with significant sequence similarity to members of cyclophilins, pathogenesis related plant proteins, lipid transfer proteins, esterases, reductases (Cadot *et al.* 2000; Flueckiger *et al.* 2002) and to actin binding proteins (Mittermann *et al.* 2005). These proteins were also identified by immunoscreening a cDNA expression library using serum from allergic patients.

In several of the reported cases the success in identifying antigens of interest is not surprising due to the fact that the antibody probes were clearly directed to specific targets using specific tissue libraries. Nevertheless, these studies prove the feasibility of identifying specific clones.

Toxic components from cnidarians often display similar primary structure and/or structural blocks (Anderluh *et al.* 2000b; Dauplais *et al.* 1997; Sher *et al.* 2005b) and therefore may have similar modes of action. It is also likely that they contain similar protein antigenicity to their homologs in related species. Consequently, the close phylogenetic relationships between chirodropids and carybdeids led to the assumption in this study that epitopes within toxic proteins could be similar and such a feature was explored.

At the beginning of this study the aim was to use antibodies to identify proteins from nematocyst crude extracts and expression libraries of *C. barnesi* and *M. kingi.* Initially, it was hoped to identify proteins homologous to those used in the production of the commercially available CSL AV. In addition, species-specific antibodies of *C. barnesi* and *M. kingi* raised in mice and sera from jellyfish sting sensitised humans were used.

2.2.2 CSL box jellyfish antivenom (CSL AV)

The only commercially-available antivenom from a toxic marine invertebrate is produced by hyper-immunization of sheep using milked venom obtained by electrical stimulation of *C. fleckeri* tentacles. Although this antivenom has been available for more than thirty years, its effectiveness to treat human envenomations has not been proven. Remarkably, there is no report of it actually saving a life - in fact three

casualties have occurred despite administration of the antivenom (Currie 2000; Currie 2003; O' Reilly *et al.* 2001; O'reilly *et al.* 2001).

Interestingly, recent experiments in animal models (piglets and rats) have shown that the effectiveness of CSL AV to counteract lethal, haemolytic, dermo-necrotic and pain-inducing effects is markedly reduced when administered crude extracts are obtained by mechanical disruption of nematocysts in comparison with milked venom and whole tentacle crude extracts (Ramasamy *et al.* 2003; Tibballs *et al.* 1998). Thus, CSL AV possibly lacks medically significant constituents (Endean & Sizemore 1988), further CSL AV is unable to neutralize the effects seen in Irukandji envenomations (Fenner *et al.* 1986).

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Collection of samples

C. barnesi samples were collected at Palm Cove, North Queensland, Australia. *M. kingi* specimens were obtained north of the outer Great Barrier Reef off Port Douglas, North Queensland (see map in figure 1.4), during the summers of 2003 to 2005. Jellyfish were either placed in sea water at 4ºC (for nematocyst isolation) or snap frozen, transported to the laboratory and stored at -80^{o}C until required.

2.3.2 Antibody probes

CSL *C. fleckeri* antivenom was produced by the Commonwealth Serum Laboratories, Australia. Human sera were obtained from four individuals with histories of jellyfish envenomations. *C. barnesi* and *M. kingi*-specific antibodies were raised by subcutaneous injection of nematocyst crude extracts (12.5 and 9.1 µg, respectively) mixed and emulsified with complete Freund's adjuvant, followed by two booster injections of the same concentration mixed with incomplete Freund's adjuvant at intervals of two weeks. Rabbit polyclonal antibodies specific for *C. fleckeri* 43/45 kDa toxins (CfTX-1 and 2) were used as well as *C. barnesi* antibodies raised in rabbit. Antibodies were stored at -20ºC or at 4ºC until used.

2.3.3 Preparation of venom extracts

Nematocysts were obtained from jellyfish using the autolysis method described by Bloom *et al.* (1998), and cellular debris was subsequently removed by ultracentrifugation in Percoll according to Marchini *et al.* (2004). Packed nematocysts were placed in water containing protease inhibitors and disrupted with glass beads in a mini bead beater, until complete nematocyst discharge was confirmed by light microscopy.

2.3.4 Protein electrophoresis and Western Blot Analysis

Proteins from crude extracts were precipitated in 10% trichloroacetic acid and cold acetone followed by centrifugation at 15,000 rpm for 15 min and a last wash with cold acetone. Protein concentrations were determined at 280 nm (using BSA as a standard) and solubilized in sodium dodecylsulphate (SDS) sample buffer (Sambrook & Russell 2001), containing β-mercaptoethanol, then denatured at 95ºC for 5 minutes. Proteins were resolved by SDS (12.5%) polyacrylamide gel electrophoresis (SDS-PAGE) in a mini-Protean II, Bio-Rad (Laemmli 1970). Apparent molecular masses were determined by comparing protein mobilities with protein standards (Weber & Osborn 1969). Proteins bands were either visualized by Coomassie blue staining (Sambrook & Russell 2001) or electrotransferred (mini-Trans-Blot, Bio-Rad) to nitrocellulose membranes. Free binding sites were blocked with 5% (w/v) skim milk in PBST for 1 hr at room temperature. Western blot analysis was performed using diluted CSL AV (1:2500), human sera (1:500), *M. kingi* and *C. barnesi* polyclonal antibodies raised in mice (1:100-500) or *C. barnesi* rabbit polyclonal antibodies (1:500). Sera from a healthy subject without cnidaria envenoming episodes, normal sheep serum, mice and rabbit pre-bleedings were used as negative controls. Primary antibodies were detected using anti-rabbit, anti-sheep, anti-human or anti-mouse AP-conjugated antibodies (1:4000-5000). Antigen-antibody binding was detected with NBT/BCIP (Promega).

2.3.5 RNA isolation and cDNA library construction

Total RNA was isolated from 1 and 2 complete specimens of *C. barnesi* and *M. kingi*, respectively. Tissue was disrupted by glass pearl beads into Trizol reagent according to the manufacturer's instructions. RNA quality was assessed by agarose gel

electrophoresis (1% w/v) and the RNA concentration determined by UV spectrophotometry. mRNA was isolated using a polyATrack system as described by the manufacturer (Promega). The mRNA of *C. barnesi* was *in vitro*-amplified (SMART^{TM} mRNA amplification, Clontech) prior to library construction. For each library, double stranded complementary DNA (cDNA) was synthesized from 5 µg of poly (A)+ template. cDNA was ligated to Uni-ZAP® XR vector (Stratagene) and packaged using a Gigapack III Gold Cloning kit (Stratagene).

2.3.6 Immunoscreening of Irukandji cDNA libraries

Lambda ZAP cDNA libraries were transfected into *E. coli* XL1-Blue MRF' cells and immunoscreening experiments were performed according to the Stratagene's Picoblue protocol. Briefly, 2 x 10^3 *pfu*/plate were spread with top agarose and incubated at 37ºC for approximately 3.5 h. Recombinant protein expression was induced using Nylon Hybond discs (Amersham Biosciences, USA) impregnated with 20 mM isopropyl-1-thio-ß-galactopyranoside (IPTG), then overlaid and incubated at 37ºC for an additional 4 h.

Filters were removed and processed for immunostaining. Non-specific binding was blocked with 0.5% skim milk in TBST, pH 7.5 for 1 h. The filters were incubated with primary antibody (CSL AV 1:2500, Human sera 1:500, species specific antibodies raised in mice -1:100-500). After three washes in PBST, membranes were incubated with secondary antibody in a dilution range of 1:2,000-5,000. The antigen–antibody complexes were visualized with NBT/BCIP (Promega) or Sigma Fast 3, 3'-diaminobenzidine (DAB) tablets. Immunopositive plaques were transferred into 50 µL of SM buffer and rescreened until single reactive plaques were obtained.

2.3.7 **In vivo** *excision of immunoreactive plaques*

pBSK-phagemids were excised from lambda Uni-ZAP® XR vector and recircularized using SOLR cells, a non-suppressing strain *of E. coli*, in the presence of ExAssist ™ helper phage (Stratagene),. Transformed bacteria were plated onto agar with carbenicillin (50 µg.mL^{-1}) and plates incubated at 37ºC overnight. Bacterial colonies

were grown and pBSK- plasmid DNA purified (Qiagen). The insert lengths of cDNA clones were estimated by restriction enzyme digestion followed by agarose (1% w/v) gel electrophoresis.

2.3.8 Sequencing, clustering and sequence analysis

Plasmid DNA from immunoreactive cDNA clones was sequenced from both ends (Macrogen, Korea). Sequence analysis and manipulations were performed using Sequencher^{TM} 4.2.2 (Gene Codes Corp.) and MacVector® 9.0 (MacVector Inc.).

Vector and adaptor sequence regions were removed from the initial dataset. A BLASTN (Altschul *et al.* 1997) comparison against the NCBI non-redundant database was conducted to identify contaminating *E. coli* DNA and M13 DNA sequences. Clones with BLAST probability scores of e^{-50} or less were eliminated from subsequent analysis.

Sequences were assembled using SeqMan II software (Lasergene) and the assemblies analysed using a variety of web-based and stand alone tools. The major open reading frames (ORFs) for consensus sequences were predicted using a reliable ORF predictor (https://fungalgenome.concordia.ca), with a default threshold of 1 x 10^{-5}. Similarity searches and comparative gene alignment analysis of both the consensus nucleotide and deduced amino acid sequences were performed using BLASTN, BLASTX and BLASTP algorithms (Altshchul *et al.* 1997) via the NCBI BLAST web server (www.ncbi.nlm.nih.gov/BLAST) or BLAST Client (blastcl3). Sequences that gave the best scores were compared using the alignment program ClustalW at the EBI web site (http://www.ebi.ac.uk/clustalw/).

Amino acid sequences were analyzed using SignalP 3.0 (Bendtsen *et al.* 2004a) and SecretomeP 2.0 (Bendtsen *et al.* 2004b) servers to predict the absence/presence and location of signal peptide cleavage sites, defined by a Neural Networks score threshold of 0.5. Codon usage, amino acid frequencies and CG content were calculated in General Codon Usage Analysis v1.0 software (McInerney 1998).

2.4 RESULT AND DISCUSSION

2.4.1 Partial characterization of native jellyfish venom proteins

Isolation, number, relative molecular mass and antigenicity of peptides and polypeptides confined to jellyfish nematocysts, in particular those of Irukandji jellyfish, were of great initial interest in this research work. However, these studies were hampered by the fact that Irukandji jellyfish samples were difficult to collect due to their small size and low numbers.

2.4.2 Nematocyst extract preparations

The isolation of *C. barnesi*, *M. kingi* and *C. fleckeri* nematocysts from jellyfish tissue was achieved using a previously described method (Bloom *et al.* 1998) and endorsed by several groups (Brinkman & Burnell 2007; Carrette & Seymour 2004; Ramasamy *et al.* 2004; Winter *et al.* 2007). The method was adapted for the small amount of samples and nematocysts were not freeze-dried. Remaining cellular debris was separated from nematocysts by ultra centrifugation in a discontinuous gradient of Percoll (Marchini *et al.* 2004). Nematocyst extracts were best achieved by bead mill homogenisation (Carrette & Seymour 2004; Estrada-Muñoz 2001), under cold conditions, at pH 7 and in conjunction with a cocktail of protease inhibitors: 0.1 mM of EDTA, 10μM of PMSF, 1 μM leupeptin and 1 μM pepstatin. It was noticed that released proteins adhered to the glass beads used for nematocyst disruption. Due to the small amount of starting material, proteins needed to be recovered by washing the glass beads twice with nematocyst extraction buffer. The diluted crude extracts were then concentrated by precipitation with TCA. Importantly, no protein degradation was detected and as a result, clear and mostly well resolved venom profiles were obtained (see Figure 2.1).

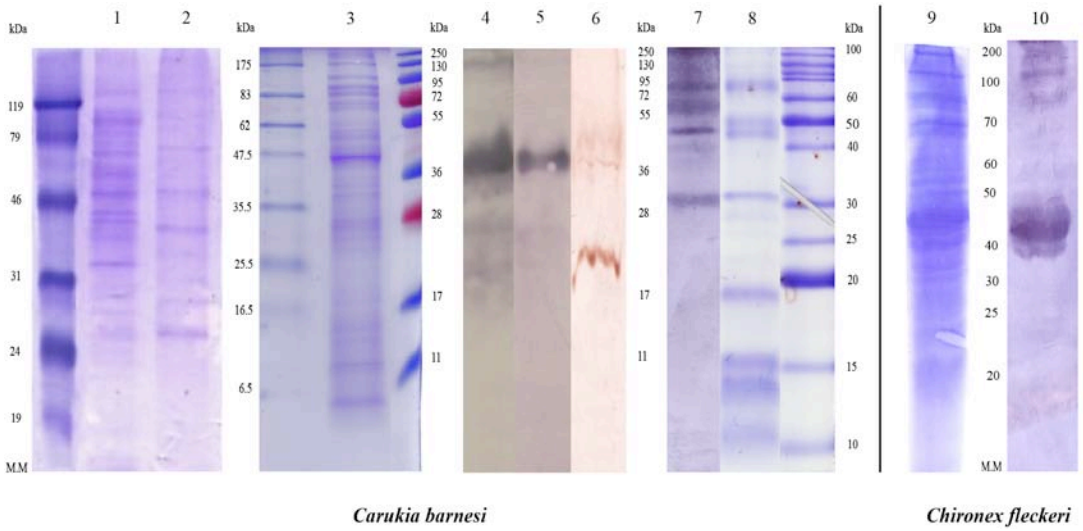


Figure 2. 1 SDS-PAGE profiles and western blot analyses of *C. barnesi* and *C. fleckeri* nematocyst extracts

Proteins were separated on 12.5% (w/v) polyacrylamide gels and western blot analysis performed. Lanes 1-6 contain *C. barnesi* samples. Lane 1 and 2: tentacular and umbrellar venom profiles, respectively; Lane 3: extract of both types of nematocysts; Lanes 4 and 5: western blot analysis of *C. barnesi* extracts tested for cross reactivity using *C. barnesi* and *M. kingi* polyclonal antibodies, respectively, raised in mice; Lane 6: cross reactivity *C. barnesi* proteins with polyclonal antibodies specific for *C. fleckeri* 43/45 kDa toxins (CfTX-1 and 2) raised in rabbit; Lane 8 *C. barnesi* nematocyst crude extract on 12.5 % SDS-PAGE and Lane 7 western blot using CSL AV, respectively; Lanes 9 and 10: *C. fleckeri* venom profile and western blot using CSL AV.

2.4.3 Venom protein profiles

Microscopic examination of *C. barnesi* nematocysts confirmed earlier observations of Southcott (1967) that two different types of nematocysts exist - spherical anisorhizas, clustered in red-like warts at the umbrella and microbasic mastigophores in the tentacles. During the course of this research we were interested in determining if there were any differences in the venom profiles of the two types of nematocysts. For that reason, three distinct venom preparations were obtained, corresponding to proteins specifically isolated from either umbrella or tentacles and the third, used for all further experiments, was isolated from both umbrella and tentacle nematocysts (whole nematocyst extract). Fractionation of toxins was accomplished using a 12.5 % (w/v) denaturing SDS-PAGE gel; the resulting venom profiles to the three extracts obtained is given in Figure 2.1: Lane 1 (umbrellar nematocytes), lane 2 (from tentacular

nematocysts) and lanes 3/7 correspond to whole nematocyst extract). Venom profile reproducibility analyses assessed by analysing the profile images with the software ImageJ64 (Data not show).

Tentacle nematocyst proteins of *C. barnesi* (Figure 2.1 Lane 1) were resolved in an apparent molecular range of 20 to 120 kDa. Molecular weight estimations and imaging analysis confirmed the presence of eleven well-resolved proteins with major bands corresponding to 47, 43, 45 and 32 kDa, followed by minor bands at 120, 90, 60 and 28 kDa. Umbrella nematocyst extract (Figure 2.1 Lane 2) of less than 4 μg total protein per well revealed eight proteins bands between 24 and 120 kDa, with major bands at 47, 45, 76-78 and 26 kDa. Separation of the proteins from *C. barnesi* whole venom (Figure 2.1 Lane 3) showed roughly 15 proteins ranging in size from 5 to 250 kDa. The profile was clearly dominated by a tight doublet seen at 46 kDa that appeared as a single band but resolved into two bands with the slightly larger band more predominant.

M. kingi possesses umbrella spherical isorhizas and tentacular microbasic mastigophore nematocysts (Gershwin 2006b). The whole crude extract was prepared with a single specimen used to isolate both types of nematocysts. Unfortunately, the scarce amount of *M. kingi* samples prevented the elucidation of the venom profile of this species. The extracted protein was used to assess the specific and overall antigenicity of the venom proteins.

In contrast with *C. barnesi* and *M. kingi*, *C. fleckeri* does not have umbrella nematocysts but instead has six distinct types of tentacular nematocysts: type 3 banana microbasic mastigophores, large oval rhopaloids, small sub-spherical rhopaloids, small rod-shaped isorhizas, and ovoid isorhizas with a spiraled tubule (Gershwin 2006b), all of which were used for whole nematocyst preparation. A profile of *C. fleckeri* nematocyst proteins was obtained that was consistent with that reported previously (Brinkman & Burnell 2007). In this research, *C. fleckeri* samples were of crucial importance as they were consistently more accessible than those of Irukandji and consequently were used not only to shape protocols but also as positive controls in venom profiles and cross reactivity studies.

The denaturing SDS-PAGE profiles revealed that there are differences not only between the proteins found in different types of nematocysts of *C. barnesi* (Figure 2.1 Lanes 1 and 2) but also that there are differences between the nematocyst proteins of *C. barnesi* and *C. fleckeri.*

It is clear that there are differences between the proteins present in *C. barnesi* nematocysts isolated from the umbrella and those isolated from the tentacles; these differences have been reported previously (Wiltshire *et al.* 2000; Underwood & Seymour 2007). It is also interesting to note that differences in protein components of nematocysts isolated from tentacle tissue of specimens from two distinct age groups have also been reported (Underwood & Seymour 2007). These differences are due to a shift in prey preferences during ontogeny, and similar changes in venom composition during ontogenetic development have been also detected in other metazoans such as snakes (Mackessy *et al.* 2006) and spiders (Herzig *et al.* 2004).

Discrepancies were observed between the banding patterns of nematocyst proteins resolved in this study and those reported previously (Underwood & Seymour 2007; Wiltshire *et al.* 2000). These discrepancies may have arisen from methodological differences such as nematocyst extraction protocols, extraction from reconstituted freeze-dried nematocysts rather than from fresh preparations, lack of protease inhibitors and subsequent protein denaturation, and poor resolution of proteins by gel electrophoresis. Notwithstanding these differences, significantly fewer protein bands were detected in venom protein profiles of both *C. barnesi* and *C. fleckeri* (Brinkman & Burnell 2007) compared with earlier studies (Underwood & Seymour 2007).

It is significant that despite heterogeneity between nematocyst protein profiles of the two difference species, there are several proteins that are common to both species, particularly a doublet around 40-46 kDa observed for both *C. barnesi* and *C. fleckeri.*

2.4.4 Antigenicity of nematocyst extracts

Mice polyclonal antisera raised against *C. barnesi* or *M. kingi* nematocyst extracts and rabbit polyclonal antibodies specific for *C. fleckeri* 43/45 kDa toxins (CfTX-1 and 2) were tested for their cross-reactivity with proteins extracted from *C. barnesi*, nematocysts. Western blot analyses (See Figure 2.1 Strips 4, 5 and 6) showed that both *C. barnesi* and *M. kingi* species-specific antibodies reacted almost equally with proteins present in the nematocyst extracts. The strongest reaction was observed for a protein at approximately 46 kDa (an unresolved protein doublet). In light of these results, it is likely that the 45-46 kDa doublet proteins of both *C. barnesi* and *M. kingi* nematocyst proteins were immunologically-dominant antigens. This may have been due to their higher abundance in the sample used to immunize the mice but may also have been due to the possibility that these proteins contain more robust exposed antigenic epitopes, hence eliciting better immunological response in immunized mice.

In Figure 2.1 Strip 6, a genuine doublet at around 45-46 kDa in the *C. barnesi* nematocyst proteins appeared to be cross-reactive with the antibodies against *C. fleckeri* cytolytic toxins CfTX-1 and -2 (Brinkman & Burnell 2007), suggesting homology to these proteins, as well as reported homologs in other species of cubomedusas including, *C. alata* CaTX-A, *C. rastoni* CrTXs and *C. quadrigatus* CqTX-A (Nagai 2003). Of concern was the presence of a strongly antigenic smaller protein of approx 20 kDa against CfTX-1 and -2 specific antibodies. Importantly, in Brinkman & Burnell (ABC News) these antibodies bound exclusively to the corresponding specific antigen. This result suggests that there may have been proteolysis of the 45-46 kDa protein or antigenicity to a naturally occurring smaller protein.

For comparison, *C. barnesi* and *C. fleckeri* nematocyst proteins were run and exposed to the CSL AV (see Figure 2.1 strip 7/8 and compare with 10/9). In the case of *C. barnesi*, protein bands of 80, 55, 30 kDa were observed and a tight doublet of 43 - 45 kDa was common for both species. The reactive proteins were genuinely specific to the CSL antivenom in that normal sheep serum did not react with any of the jellyfish extracts used. Moreover, it is clear that the *C. fleckeri* 43 and 45 kDa proteins are

immuno-dominant components in CSL AV. These results are consistent with previous studies (Brinkman & Burnell 2008).

Figure 2.2 shows dot blots of venom proteins from *C. barnesi*, *M. kingi* and *C. fleckeri* on nitrocellulose membranes that were screened for immuno-reactivity towards Irukandji species-specific antibodies, human sera and CSL AV. The dot blot analyses highlighted differential antigenicity of jellyfish proteins versus the different sources of antibodies. CSL AV was reactive not only to *C. fleckeri* extracts but also to Irukandji jellyfish nematocyst proteins; this is consistent with previous results (Wiltshire *et al.* 2000), experimental results in which CSL AV reacted with jellyfish extracts of *C. barnesi* and also with total proteins extracted from other medusae. These results indicate that despite jellyfish venoms having different components, toxicity and potency, there are shared epitopes eliciting specific reactivity. Nevertheless, this cross reactivity seems to have an evolutionary rather than a clinical significance due to the fact that CSL AV is unable to neutralize the effects of *C. barnesi* envenomation in humans (Fenner *et al.* 1986b).

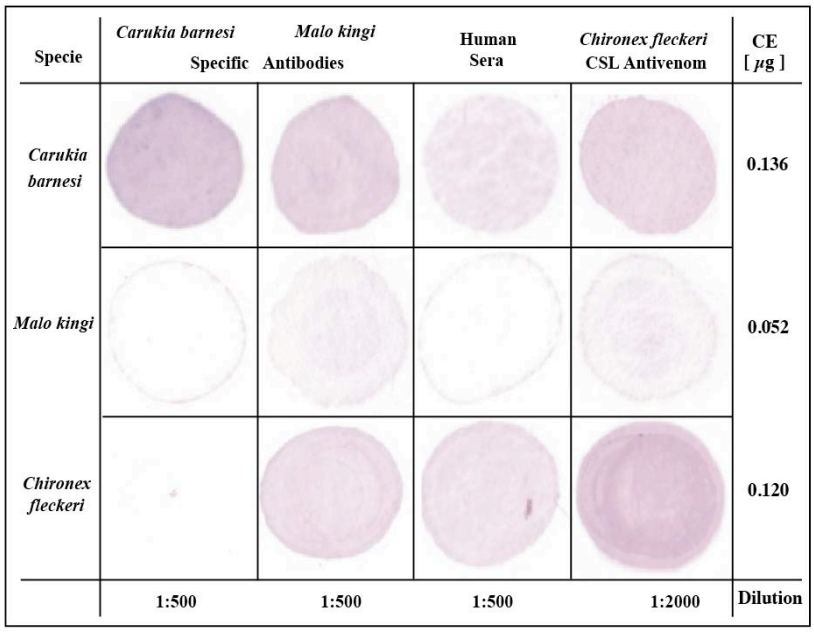


Figure 2.2 Differential cross-reactivity among jellyfish crude extracts.

Dot blot analysis of jellyfish nematocyst extracts of three cubozoans, *C. barnesi*, *M. kingi* and *C. fleckeri* tested for immune reactivity using four probes: *C. barnesi* and *M. kingi* specific polyclonal antibodies raised in mice, human sera and CSL AV. The optimised dilution and the protein concentrations used are indicated. CE, refers to nematocytes crude extracts.

Interestingly, reactivity among the four human sera was found to be different in potency of cross-reaction against jellyfish extracts (data not displayed). This could possibly be explained by a boosting effect caused by repeated jellyfish stings in already sensitised individuals. Although the major elicitors of IgE response in subjects are unknown, it is clear that human seras are reactive against jellyfish extracts and that one of the sera caused greater reactivity and therefore was used in further experiments.

The mice pre-bleeding sera, normal sheep sera and serum from a non-stung (naïve) human were not cross reactive with any jellyfish proteins whereas the primary probes clearly reacted against the three jellyfish extracts, demonstrating that genuine binding of experimental antibodies had taken place on the dot blot.

2.4.5 Molecular aspects of C. barnesi *and* M. kingi *cDNA libraries*

The *C. barnesi* cDNA library originally had a titre of 3.3 x 10^7 *pfu*.ml^{-1} and was amplified to 2 x 10^{12} *pfu*.ml^{-1}. The average insert size was determined by restriction endonuclease digestion assays of more than 100 cDNAs to be 1,728 bp. Inserts larger than 3,000 bp were common. The clonal representation in this library was homogeneous, with particularly long inserts, long polyadenylated regions, intrusion of chimerical clones and unusual adenine repeat regions in some cDNA clones observed. It is believed these arose from the *in vitro* mRNA amplification process and perhaps genomic DNA contamination. Nevertheless, unusual clones are not over-represented.

The *M. kingi* cDNA library was non-normalized and not amplified with a titre of 9 x 10^{10} *pfu*.ml^{-1} and an average insert size of (1600 bp) mainly in the range of 0.6 to 2.5 kb. Clonal representation was homogeneous.

2.4.6 Codon usage in C. barnesi *and* M. kingi

Calculation of codon usage was performed using translated sequences of sequenced clones from the cDNA libraries. *C. barnesi* codon usage was calculated for 50 clusters (203 cDNA reads) and *M. kingi* codon usage was calculated for 1,039 clusters (1,466 cDNA reads). In total 10,866 and 250,503 triplets of *C. barnesi* and *M. kingi*, respectively, were used and the usage is summarized in Table 2.1, indicating that in cubomedusae bias exists for certain codons.

Table 2. 1 Accumulative codon usage of both *C. barnesi* and *M. kingi*

In here, N refers to the calculated number of times a particular codon is observed in a gene or set of genes. RSCU refers to the Relative Synonymous Codon Usage values for a given data set. This RSCU value represents the number of times that a codon would be observed in the absence of any codon usage preference. The RSCU would is set to be 1. This mean that when a codon is used less frequently, it will have a value less than 1 and vice versa for a codon that is used more frequently than expected (McInerney 1998).

	Carukia barnesi					*Malo kingi*				
AA	Codon	N	RSCU	AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	410	(1.29)	Ser	UCU	212	(1.21)	Phe	UUU	10793	(1.35)	Ser	UCU	4487	(1.27)
	UUC	226	(0.71)		UCC	162	(0.93)		UUC	5235	(0.65)		UCC	2826	(0.80)
Leu	UUA	231	(1.25)		UCA	216	(1.24)	Leu	UUA	5948	(1.38)		UCA	4778	(1.36)
	UUG	241	(1.31)		UCG	116	(0.66)		UUG	5848	(1.36)		UCG	1978	(0.56)
Tyr	UAU	200	(1.23)	Cys	UGU	207	(1.08)	Tyr	UAU	5734	(1.23)	Cys	UGU	5140	(1.21)
	UAC	126	(0.77)		UGC	177	(0.92)		UAC	3607	(0.77)		UGC	3380	(0.79)
End	UAA	216	(0.00)	End	UGA	221	(0.00)	End	UAA	5843	(0.00)	End	UGA	4900	(0.00)
End	UAG	113	(0.00)	Trp	UGG	145	(1.00)	End	UAG	2965	(0.00)	Trp	UGG	3569	(1.00)
Leu	CUU	239	(1.30)	Pro	CCU	116	(1.09)	Leu	CUU	5262	(1.22)	Pro	CCU	2699	(1.20)
	CUC	138	(0.75)		CCC	70	(0.66)		CUC	2537	(0.59)		CCC	1745	(0.77)
	CUA	125	(0.68)		CCA	156	(1.46)		CUA	2813	(0.65)		CCA	3347	(1.49)
	CUG	133	(0.72)		CCG	85	(0.80)		CUG	3448	(0.80)		CCG	1218	(0.54)
His	CAU	190	(1.27)	Arg	CGU	89	(0.82)	His	CAU	4538	(1.21)	Arg	CGU	1799	(0.75)
	CAC	109	(0.73)		CGC	59	(0.54)		CAC	2968	(0.79)		CGC	1109	(0.47)
Gln	CAA	251	(1.16)		CGA	105	(0.96)	Gln	CAA	5712	(1.22)		CGA	1960	(0.82)
	CAG	181	(0.84)		CGG	59	(0.54)		CAG	3635	(0.78)		CGG	1154	(0.48)
Ile	AUU	265	(1.25)	Thr	ACU	200	(1.16)	Ile	AUU	7399	(1.30)	Thr	ACU	3882	(1.15)
	AUC	181	(0.86)		ACC	136	(0.79)		AUC	3885	(0.68)		ACC	2765	(0.82)
	AUA	189	(0.89)		ACA	264	(1.53)		AUA	5742	(1.01)		ACA	4961	(1.48)
Met	AUG	177	(1.00)		ACG	89	(0.52)	Met	AUG	4758	(1.00)		ACG	1843	(0.55)
Asn	AAU	288	(1.20)	Ser	AGU	162	(0.93)	Asn	AAU	7335	(1.24)	Ser	AGU	4063	(1.15)
	AAC	193	(0.80)		AGC	179	(1.03)		AAC	4496	(0.76)		AGC	2995	(0.85)
Lys	AAA	409	(1.26)	Arg	AGA	235	(2.15)	Lys	AAA	11367	(1.34)	Arg	AGA	5211	(2.19)
	AAG	239	(0.74)		AGG	108	(0.99)		AAG	5606	(0.66)		AGG	3068	(1.29)
Val	GUU	192	(1.38)	Ala	GCU	157	(1.25)	Val	GUU	4608	(1.34)	Ala	GCU	2921	(1.24)
	GUC	131	(0.94)		GCC	112	(0.89)		GUC	2539	(0.74)		GCC	1869	(0.79)
	GUA	106	(0.76)		GCA	162	(1.29)		GUA	3462	(1.01)		GCA	3448	(1.47)
	GUG	127	(0.91)		GCG	70	(0.56)		GUG	3142	(0.91)		GCG	1176	(0.50)
Asp	GAU	159	(1.03)	Gly	GGU	113	(0.97)	Asp	GAU	4049	(1.21)	Gly	GGU	2748	(1.14)
	GAC	151	(0.97)		GGC	109	(0.94)		GAC	2668	(0.79)		GGC	1905	(0.79)
Glu	GAA	262	(1.32)		GGA	155	(1.33)	Glu	GAA	5774	(1.34)		GGA	3277	(1.36)
	GAG	134	(0.68)		GGG	88	(0.76)		GAG	2865	(0.66)		GGG	1701	(0.71)

2.4.7 Identification of C. barnesi *and* M. kingi *cDNA antigenic clones*

It is clear that all of the jellyfish polyclonal antibodies showed some degree of cross reactivity with jellyfish proteins isolated from nematocysts (See dot blot Figure 2.2) and this fact lead us to screen the Irukandji expression libraries using these antibodies.

Antigenic clones (39 in total) were isolated and characterized from Irukandji cDNA libraries by immunoscreening using CSL AV and *C. barnesi* whole polyclonal specific antibodies raised in rabbits. No antigenic clones resulted from screening the *C. barnesi* cDNA library using human sera or *C. barnesi*/*M. kingi* polyclonal antibodies of nematocyst proteins raised in mice. In view of that fact, no further attempts to screen *M. kingi* cDNA libraries with those probes were made.

Table 2.2 indicates that immunoreactive cDNA clones encoded a wide variety of proteins with a diverse range of functions. Predominantly these were house-keeping genes that are not related to defence mechanism processes. No appreciable similarity was observed with any toxin currently known in cnidarians or in any other metazoan animal.

Table 2. 2 Immunoreactive clones clustered according to apparent function

Funtion	Immunoreactive clone ID and annotation
Structural	Mk-CSL AV-1-actin
	Mk-CSL AV-4-non muscle myosin heavy chain
	CbRS-9- Minicollagen component of nematocyst wall
	Cb-CSL AV-5-Thypedin
	Cb-CSL AV-4- Clathrin heavy chain
Transcription	Cb-CSL AV-2-Ribosomal protein L10/QM-like
	CbRS-7-40S Ribosomal protein S4
	CbRS-11-Ribosomal protein L16p/L10e
	Mk-CSL AV-6-Ribosomal protein L21
	Mk-CSL AV-7-Ribosomal protein L35Ae
	CbRS-8-Eukaryotic translation elongation factor 2
	Mk-CSL AV-2-Elongation factor 1 alpha
	Mk-CSL AV-3 1624TPA: gag-pol polyprotein
Stress response	Mk-CSL AV-11- Trypsin-like serine protease (Mk-ShK-TSP-2)
Outer membrane interaction	Cb-CSL AV-1- Myosin-like antigen
Enzymes	Cb-CSL AV-9-Complement component C3-like cDNA
	Cb-CSL AV-8-similar to Alpha-N-acetylglucosaminidase
Molecular chaperones	Mk-CSL AV-5-Cyclophilin A, Cb-CSL AV-3-Heat shock protein 70,
	Mk-CSL AV-8-Similar to HSP90-like ATPase. 90
Metal binding	CbRS-10-Ferritin
	Cb-CSL AV-10-Zfing gene -Similar to Zinc finger
Novel	Cb-CSL AV-6, Cb-CSL AV-7, Cb-CSL AV-12, CbRS-3,
	CbRS-4, CbRS-5, Mk-CSL AV-9, Mk-CSL AV-10

cDNA clones were characterized by their partial or complete DNA sequence. Relevant features *e.g.*, ribosomal binding sites, stop/start codon, 5'- and 3'-untranslated regions (UTRs), ORFs, nucleotide repeat regions, and splicing sites were identified. In addition, peculiar features of each cDNA clone and the conceptual primary structure features such as domain architecture repetitions, functional category, and involvement in secretory pathways were identified. Primary structure was also analyzed for potential antigenicity using the protein module of MacVector to predict potential antigenic sites and regions having a high antigenic index. The characterization of reactive cDNA clones resulting from the three immunoscreening experiments are summarized in Tables 2.3, 2.4 (Cb-CSL AV), Table 2.5 (Cb-RHS) and Table 2.6 (Mk-CSL AV). Complete ORFs for some of the clones were not obtained. This subset of cDNAs is not likely to express a functional fusion protein. Collectively these data suggest that various clones may have displayed non-specific reactivity.

2.4.8 CSL AV as probe against Irukandji cDNA libraries

Immunoscreening of *C. barnesi* and *M. kingi* cDNA libraries yielded 26 reactive clones which did not appear to be involved in envenomation and the resulting clones resembled random sequencing (ESTs) rather than specific targeting.

It is relevant to mention the Yang and colleagues (2003a) report of an EST analysis from the scyphozoan, *C. capillata*. This study yielded 1,016 unigenes. Within the ten most abundant transcripts were five different ribosomal proteins (l30, l35, S-29, L32 and S20), ferritin, elongation factor 1A, cytochrome C, the cytoskeletal protein thrombospodin, and a cyclophilin. The study showed a very well represented data set, which indicated clearly a high expression level of those genes. The vast majority of the clones described were housekeeping proteins and the representation of the *C. capillata* data and the Cubozoa immunoreactive clones were similar. This indicated that Irukandji reactive clones as a whole resembled random sequencing and were not selected by antibody binding. However, it is relevant to point out that there are several

proteins whose related homologs in other animals act as allergens. This is the case of ribosomal proteins (RP) and cyclophilin A. In Table 2.6 three different cDNA clones comprise complete coding regions for RP; Mk-CSL AV-6, Mk-CSL AV-7 (L21) and cyclophilin A (Mk-CSL AV-5).

Table 2. 3 *C. barnesi* and *M. kingi* gene phylogenetic comparisons results (BLASTP) of immunologically-reactive cDNA clones.

This table summarizes sequence similarities, gene descriptions, probability of occurrence by chance and gene class. Rows shaded in green correspond to cDNA clones immunoreactive towards CSL *C. fleckeri* antivenom. Shaded in blue are immunoreactive to *C. barnesi* polyclonal antibodies raised in mice. Gene classification is disclosed as in the scyphomedusa *C. capillata* report only simplified (Yang *et al.* 2003a) and is divided into: (**A**) housekeeping genes, (**B**) cell-cell communication, (**C**) transcription factors and (**D**) miscellaneous (not enough information to classify, or no significant similarities to known proteins).

Cluster ID	Data base entry name	Accession Number	Species	Score	e.	Gene Class
Cb-CSL AV-1	Non-muscle myosin/ myosin-like antigen	gb|AAB50272.1|	*Onchocerca volvulus*	106	4e-21	(**A**) Cytoskeleton and membrane proteins
Cb-CSL AV-2	Ribosomal protein L10/QM-like protein	ref|XP_787854.1|	*Strongylocentrotus*	258	1e-67	(**A**) Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
Cb-CSL AV-3	Heat shock protein 70	gb|AAO38780.1|	*Chlamys farreri*	393	1e-108	(**A**) Stress response, detoxification and cell defense proteins
Cb-CSL AV-4	Clathrin assembly protein	ref|NP_001073586.1|	*Gallus gallus*	468	1e-130	(**A**) Protein degradation and processing, proteases
Cb-CSL AV-5	Thypedin	gb|AAW82079.1|	*Hydra vulgaris*	349	1e-94	(**A**) cytoskeleton organization, membrane proteins and biogenesis
Cb-CSL AV-6	Nodulin- Aquaporin-like	gb|AAA02947.1|	*Glycine max (soybean)*	42	0.036	(**D**) Not enough information to classify
Cb-CSL AV-7	Predicted protein	gb|EDO40849.1|	*N. vectensis*	30	3.6	(**D**) Not enough information to classify, No significant similarities to known proteins
Cb-CSL AV-8	Alpha-N-acetylglucosaminidase	ref|XP_414709.2|	*Gallus gallus*	127	2e-36	(**A**) Intermediary synthesis and catabolism enzymes
Cb-CSL AV-9	Alpha-1-macroglobulin	sp|Q63041|A1M_RAT	*Tetraodon nigroviridis*	177	1e-42	Endopeptidase inhibitor
Cb-CSL AV-	Predicted protein	ref|XP_001640900.1|	*N. vectensis*	98	2e-21	(**A**) Transportation and binding proteins for ions and other small molecules
Cb-CSL AV-	Apolipophorin precursor	CAB51918.2	*Locusta migratoria*	98.6	6e-19	
Cb-CSL AV-	Predicted protein	gb|EDQ93056.1|	*Monosiga brevicollis MX1*	138	5e-31	**(A)** Intermediary synthesis and catabolism enzymes
Cb-CSL AV-	Probable Cytochrome P450 like_TBP	pir||T02955	*Zea mays*	129	6e-28	(**A**) Stress response, detoxification and cell defense proteins
Cb-CSL AV-	RACK -Receptor of protein Kinase	dbj|BAE93065.1|	*Petromyzon marinus*	174	2e-42	(**A**) Cell replication, histones, cyclins and allied kinases, DNA polymerases
CbRS-1	Predicted protein	ref|XP_001634593.1|	*N. vectensis*	518	1e-145	(**A**) Amino acid metabolism
CbRS-2	Cell adhesion protein. fasciclin I-like protein	gb|AAM28437.1|	*Anthopleura elegantissima*	159	5e-37	(**B**) Intracellular signal transduction pathway molecules including kinases and signal
CbRS-3	Type I secretion target repeat protein	ref|ZP_01054714.1|	*Roseobacter sp*	40	0.28	
CbRS-4	***** No hits found ******					(**D**) No significant similarities to known proteins
CbRS-5	Putative multidrug ABC transporter	ref|YP_053505.1|	*Mesoplasma floru*	35	5.3	(**D**) No significant similarities
CbRS-6	Mitochondrial ATP synthase F0 complex subunit	gb|ABF22409.1|	*Takifugu rubripe*	88	6e-16	
CbRS-7	Predicted protein	ref|XP_001625647.1	*N. vectensis*	345	1e-93	(**A**) Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
CbRS-8	Eukaryotic translation elongation factor 2	gb|AAP49571.1|	*Aurelia aurita*	232	1e-61	(**A**) protein synthesis elongation
CbRS-9	mini-collagen	emb|CAA43379.1|	*Hydra sp.*	43	8e-05	
CbRS-10	Ferritin	ref|XP_001623678.1|	*N. vectensis*	240	5e-62	(**A**) Transportation and binding proteins for ions and other small molecules
CbRS-11	Ribosomal protein L16p/L10e	ref|XP_001640327.1|	*N. vectensis*	119	8e-26	(**A**) Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
CbRS-12	Flap endonuclease-1-Predicted protein	ref|XP_001651504.1	*Aedes aegypti*	385	9e-106	(**D**) Unable to classify
CbRS-13	Vitellogenin	gb|AAL01527.1|	*Larus argentatu*	75	7e-12	(**A**) Transportation and binding proteins for proteins and other macromolecules, lipid
CbRS-14	Similar to dihydrolipoamide S-acetyltransferase	ref|XP_001626116.1|	*N. vectensis*	213	8.e-54	(**A**) Intermediary synthesis and catabolism enzymes cell metabolism, (low prevalent
Mk-CSL AV-1	Actin	sp|P17126|ACT_HYDAT	*Hydra vulgaris*	560	1e-158	(**A**) Cytoskeleton and membrane proteins
Mk-CSL AV-2	Elongation factor 1 alpha	emb|CAE45763.1|	*Axinella verrucosa*	571	1e-161	(**A**) Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
Mk-CSL AV-4	Similar to non muscle myosin heavy chain	ref|XP_683046.2|	*Danio rerio*	156	1e-36	(**A**) Cytoskeleton and membrane proteins
Mk-CSL AV-3	TPA: gag-pol polyprotein	tpe|CAJ00252.1|	*Schistosoma mansoni*	102	1e-19	
Mk-CSL AV-5	Cyclophilin A	gb|AAR11779.1|	*Chlamys farreri*	203	6-51	(**A**) Protein degradation and processing, proteases
Mk-CSL AV-6	Ribosomal protein L21	ref|XP_001631045.1|	*N. vectensis*	216	5e-55	(**A**) Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
Mk-CSL AV-7	Ribosomal protein L35Ae	gb|ABC25032.1|	*Hydra vulgaris*	142	9e-33	**(A)** Protein synthesis co-factors, tRNA synthetases, ribosomal proteins
Mk-CSL AV-8	Predicted protein similar to HSP90-like ATPase.	ref|XP_001637729.1|	*N. vectensis*	450	1e-124	**(A)** Transportation and binding proteins for proteins and other macromolecules
Mk-CSL AV-9	Predicted protein	ref|XP_001627127.1|	*N. vectensis*	74	1e-11	Novel Gene, not classified yet (refer to gene characterization. Table 9)
Mk-CSL AV-	Predicted protein Trypsin-like serine protease	ref|XP_001640431.1|	*N. vectensis*	144	3e-33	

Table 2. 4 Characterization of *C. barnesi* cDNA clones shown to be immunoreactive towards CSL *C. fleckeri* antivenom (CSL AV)

Rows shaded in yellow correspond to either full-length cDNA clones or comprising complete encoding regions. Pale green rows correspond to partial cDNAs (lacking either or both 5'/ 3' region) and hence partial protein, gaps and cDNA inserts were estimated by digestion assay using either combination of restriction enzymes within the pBSK- polylinker. Red asterisk * = proteins that undertake a classic or non-classical secretion pathway, as predicted using SignalP (Bendtsen *et al.* 2004a) and SecretomeP 2,(Bendtsen *et al.* 2004b) and obtaining neural network values exceeding the normal threshold of 0.5. IDs underlined are clones with significant predicted antigenic epitopes. Conceptual translations were searched in diverse data bases: Simple Modular Architecture Tool (SMART) (Schultz *et al.* 1998), Protein Family data base (Pfam) and Interpro entries were used to enrich annotations. CnidBase. Putative O-glycosylated and N-glycosylated sites were predicted using NetPhos 2.0 and NetOGlyc 3.1 Server (Blom *et al.* 1999; Julenius *et al.* 2005). Rows shaded in orange correspond to partial cDNAs but encoding a complete post-translational product (predicted by gene similarity).

Clone ID	Length	Accession	Annotation	Relevant details about the cDNA clone and the encoded product.
Cb-CSL AV-1	1320	EU753860	Myosin-like antigen	cDNA clone sequenced completely. No 5' UTR and no triplet for initial methionine, 202 bp of 3' UTR, 1117 bp of coding region, no poly A tail. By homology this cDNA clone is missing around 1089 bp. The clone encodes 363 amino acids of a Myosin-related protein, according with homologs the complete product should be around 661 amino acids. The protein fragment comprises two complete AAA domains of ATPases associated with a variety of cellular activities. (e value of 1.00e-18). Homologs of this protein have been shown to be associated predominantly with hypodermis and a number of other specific membrane layers in adult parasite. Myosin-related proteins are frequently immunodominant parasite antigens and in a number of studies have been shown to confer a degree of protective immunity against the corresponding parasite. (Shang & Miller 1995). Likely to be homologs to *C. capillata*: Q9N2R3 a heat stable allergen tropomyosin.
Cb- CSL AV -2	576	EU753861	Ribosomal protein L10/QM-like protein	A truncated cDNA clone sequence completely. No 5' UTR but contains 3' UTR and poly A tail is observed. According with phylogenetic alignments at CnidBase Cb-CSL AV-22 - RL10 is missing 50 amino acids.
Cb-CSL AV-3	854	EU753862	Heat shock protein 70-Pseudogene	Truncated cDNA clone. Contains exclusively coding region. No stop codon but there is a poly-adenylated signal. No sequencing errors, presumably a fragmented pseudogene. Note. Related sequences and not pseudogene were found in the *M. kingi*: EST collection. (EST_ 723, 641 and 358)
Cb-CSL AV-4 *	1594	EU753863	Similar to Clathrin heavy chain	No complete sequence (2000 bp confirmed by restriction assay using *Xho*I/*Eco*RI, cDNA clone inverted. Search of the protein domain database SMART, detected three Clathrin heavy chain repeats from positions 22 to 181, 206 to 347and 354 to 499 with e value of 1.71e-39, 1.28e-20 and 1.38e-43 respectively. Clathrin is associated with intracellular transfer of membrane by coated vesicles. Clathrin is a single major protein (apparent MW=180,000) that forms the coat of vesicles. Peptide mapping suggests that the amino acid sequence of Clathrin is conserved, irrespective of tissue or species. Coated vesicles of different sizes are found and are constructed with variable numbers of Clathrin subunits, arranged in closed networks of hexagons and pentagons (Pearse 1976).
Cb-CSL AV-5	854	EU753864	Thypedin	Completely sequenced, comprising coding region (238 residues long), it is missing 100 residues for the protein sequence to be complete. The cDNA contains a poly A tail. Cellular component: cytoplasm, molecular function: actin binding Biological process: cytoskeleton organization and biogenesis (inferred from electronic annotation from InterPro).
Cb-CSL AV-6	1023	EU753865	Hypothetical protein	Not completely sequenced. Similar to aquaporins, these proteins contain two tandem repeats each containing three membrane-spanning domains and a pore-forming loop with the signature motif Asn-Pro-Ala (NPA). Cb-CSL AV-10-3' end display high similarity to the most C-terminal repeat of aquaporins but the sequence lacks the NPA signature because it is not a completely sequenced clone. The function of aquaporins is to facilitate the transport of water and small neutral solutes across cell membranes.
Cb-CSL AV-7	1320	EU753866	Predicted protein	Not completely sequenced, a 1500 bp fragment was confirmed by digestion using *Xho*I/*Eco*RI indicating that there is a sequence gap of approximately 200 bp. The 5' read is 591 bp long. No acceptable match similarity. No known domains detected.
Cb-CSL AV-8	1020	EU732708	Similar to Alpha-N-acetylglucosaminidase	Fully sequenced, contains 156 bp and 303 bp of 5' and 3' UTR. No poly A tail. The clone encodes a complete product that contains a signal peptide beginning at residue 1 and ending at residue 31, followed by a partial domain of exopolysaccharide biosynthesis protein that is related to N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. According to protein similarity analyses, this protein is involved in carbohydrate transport and metabolism and its gene ontology falls in intermediary synthesis and catabolism enzymes.

Table 2. 4 *Continuation* **Characterization of *C. barnesi* cDNA clones shown to be immunoreactive towards CSL *C. fleckeri* antivenom (CSL AV)**

Clone ID	Length	Accession	Final Annotation	Relevant details about the cDNA clone and the encoded product.
Cb-CSL AV-9	1215	EU732709	C3-like cDNA. Thiol ester-containing proteins	Very interesting gene. Fully sequenced cDNA clone. Comparison among closest homologs suggest that the clone is lacking at least 3510 bp (approximately 1170 amino acids) to be full length. The truncated cDNA clone does not contain 5' terminal untranslated region but has 214 bp of 3' UTR. A canonical polyadenylation signal site (AATAAAA) is observed but it lacks a poly A tail. The coding region lacks an initial methionine and the stop codon is located at position 919. The ORF is 909 bp encoding 303 amino acids. SMART detected two complete C-terminal domains: position 1 to 84 - A-macroglobulin complement component, intrinsic disorder region from 145 to 162 and from position 184 to 274 A-macroglobulin receptor. Alpha (2)-M constitutes a large family of large glycoprotein inhibitors (Xiao *et al.* 2000). These broadly specific proteinase inhibitors are major carrier proteins in serum. According to sequence homology the encoded product of Cb-CSL AV-9 TEP-Alpha-2-M-like, belongs to the MEROPS proteinase inhibitor family I39, clan IL. These protease inhibitors share several defining properties, which include (i) the ability to inhibit proteases from all catalytic classes, (ii) the presence of a 'bait region' and a thiol ester (Heralde *et al.* 2008), a similar protease inhibitory mechanism and (iv) the inactivation of the inhibitory capacity by reaction of the thiol ester with small primary amines. Alpha 2 M protease inhibitors inhibit by steric hindrance (Enghild *et al.* 1989). Most of A2M are 180 kDa and under physiological conditions often occurs as tetramers, although some exist as functional monomers or dimers, especially in phylogenetically ancient species. (Sottrup-Jensen 1987). Genes similar to Cb-CSL AV-9 are exclusively found in Metazoan.
Cb-CSL AV-10	890	EU732707	Zfing-gene (Similar to Zinc finger.)	cDNA clone sequenced completely, Kozak sequences and poly A tail observed. There is a clean ORF of 133 residues. This protein is around 15kDa, with an estimated pI= 9.32. In addition the protein contains a Furin-type cleavage site after the 25th residue. Seven putative phosphorylation sites at positions 12, 40, 43, 51, 55, 71 and 98. Functional category is predicted to be translation. This relatively small protein contains no know domains. By similarity it is related to zinc finger (Znf) domain containing proteins. Found low but significant similarities with domains such as c4 zinc finger in nuclear hormone receptors, GAL4 a Rubredoxin domain, which is involved in electron carrier activity and metal ion binding. Remarkably the only homologs sequence comes from the starlet anemone with Identities = 52/125 (41%), Predicted protein from *N. vectensis* (DS469512) VERY ANTIGENIC. Novel gene
Cb-CSL AV-11	1005	bankit1096637	Lipovitellin-1	This cDNA clone contains 822 bp of encoding region, followed by TAG at position 823-825 and the 3' UTR ending with a poly A tail. The encoded product is 275 amino acid residues long. From position 46 to 270 comprises a complete domain of unknown function (DUF1943) found in vitellin-like proteins. Calculated molecular weight of the putative protein after posttranscriptional modification= 31384.64, estimated pI = 8.15. This cDNA clone reacted against *C. fleckeri* antivenom, however it is unlikely to be a truly positive-antigenic clone. Nevertheless, I give a partial characterization of this clone due to its relevance. (DUF1943)- "Members of this family adopt a structure consisting of several large open beta-sheets. Their exact function has not yet, been determined; pfam09172. Vitellogenins (Vgs) are secreted, multi-domain proteins that undergo posttranslational modifications. The characteristic vertebrate vitellogenin is composed of an N-terminal signal peptide, and four post translationally modified vitellins (yolk proteins): Lipovitellin-1, phosvitin, lipovitellin-2 and a von Willebrant factor type D domain. (paraphrased from InterPro abstract IPR001747). However, there are about 12 different architecture combinations of the above and two other minor domains. Vitellogenins are found abundantly in the yolk of egg laying animals (Thompson & Banaszak 2002) that provide the mayor egg yolk proteins involved in metal storage and is an essential source of nutrients during early development in oviparous vertebrates and invertebrates (such as nematodes, arthropods and molluscs). There are two related products in Cnidarians (dbj|BAD74020.2| egg protein *Galaxea fascicularis*, 1439 a.a (Hayakawa H 2002) and recently released gb|EDO42586.1| predicted protein *N. vectensis*, 1488 a.a. Both, the coral and the anemone protein architectures denote the same domain pattern: after a signal peptide is followed by a Lipoprotein N-terminal domain (LPD_N), DUF1943 and a von Willebrant factor type D domain (vWF) at the C-termini, thus resembling the vertebrate vitellogenin domain signature. In contrast the fragmented protein identified in *C. barnesi*, only comprises a complete DUF1943 domain found at the C-termini rather than a vWF motif. Similar architecture is seen in the vitellogenin of the pacific oyster *Crassostrea gigas* (Q8IU34_CRAGI) and also in the homologs from *Tetraodon nigroviridis* Q4SP71_TETNG. I conclude that this Cubozoan vitellogenin-like cDNA clone encodes a partial fragment (the most C-terminal) of the Cb vitellogenin. Due to the posttranslational modifications of all Vgs homologs, it is likely that this jellyfish Vg fragment is a complete and functional egg protein that I have named "Cb-Lipovitellin-1". On the basis of protein alignments of the DUF1943 among the box jellyfish, the coral and the anemone but mainly due to the differences found in the domain architecture distribution, I presume that the Cb vitellogenin precursor is a new member of the Vg family in Cnidarians. In spite of the fact that this cDNA encodes a non-complete Vg protein precursor, it served as a stimulus to partially characterize a putative new member of this interesting protein family. A complete Cubozoa Vg precursor protein sequence, in time, may lead to the identification of a more complete evolutionary and structural relationship of Vgs within the phylum Cnidaria and other metazoans.
Cb-CSL AV-12	864	EU753867	Predicted protein	cDNA clone sequenced completely, poly A tailed. The clone encodes an ORF of 104 residues, ALDH-like protein belongs to the aldehyde dehydrogenase family. This family of dehydrogenases act on aldehyde substrates. Members use NADP as a cofactor. The family includes the following members: the prototypical members are the aldehyde dehydrogenases, succinate; pfam00171. Similarities were found with other Cnidaria: predicted protein *N. vectensis* (gb|EDO38923.1) with score of 34 an e value of 1e-31.
Cb-CSL AV-13 *[?]	1359	EU753868	No apparent coding region.	cDNA clone fully sequenced. Several stop codons breaking the ORF with a poly A tail. Homologs in *Zea mays*, *Escherichia coli* and *Saccharomyces cerevisidae* are suppose to have a cytochrome P450-like encoding cDNA containing an open reading frame of 1593 bp and encoding a protein with a molecular size of 58916. There are no propeptide cleavage sites.
Cb-CSL AV-14	362	EU753869	Similar to RACK -Receptor of protein Kinase	cDNA clone sequenced completely, no poly A tail. It encodes proteins with two WD40 repeats, found in a number of eukaryotic proteins that cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly. The first WD domain observed in this fragment starts at position 1 and ends at position 38. Terminating in WD and it is similar to the one found in guanidine nucleotide binding protein subunit 1 (Receptor of activated protein kinase C). The second motif starts at position 59 and ends in position 99; the last two residues are WQ. The WD domain often ends in WD, hence the name. Proteins with 4-16 WD repeats are implicated in functions such as signal transduction, transcription regulation, cell cycle control and apoptosis. The common function is to coordinate multi-proteins assemblies.

Table 2.5 *C. barnesi* cDNA clones immuno reactive towards *C. barnesi* specific antibodies raised in rabbit

Rows shaded in yellow correspond to either full-length cDNA clones or comprising complete encoding regions. White rows correspond to partial cDNAs (Lacking either 5' or 3') and hence partial protein (NCPP), Gaps and cDNA inserts were estimated by digestion assay using either combination of cutters within the Pbk- polylinker. Red asterisk * = Proteins that undertake a classic or non-classical secretion pathway were predicted using SignalP and SecretomeP 2 (Refs). obtaining neural network values exceeding the normal threshold of 0.5 (Bendtsen *et al.* 2004). IDs underlined are clones with significant predicted antigenic epitopes. Conceptual translations were searched in diverse data bases: Simple Modular Architecture Tool (SMART) (Schultz *et al.* 1998), Protein Family data base (Pfam) and Interpro entry were used to rich annotations. CnidaBase . Putative O-glycosylated and N-glycosylated sites were predicted using NetPhos 2.0 and NetOGlyc 3.1 Server (Blom *et al.* 1999; Julenius *et al.* 2005)

Clone ID	Length	Accession	Final Annotation	Final Annotation Relevant details about the cDNA clone and the encoded product.
CbRS-1	1449	EU878241	cystathionine	Full length clone encoding cystathionine beta-lyase/cystathionine gamma-synthase. It is involved in Amino acid transport and metabolism. Form position 12 to 389 contains a Cys_Met_Meta_PP""Cys/Met metabolism PLP-dependent enzyme (pfam01053). This family includes enzymes involved in cysteine and methionine metabolism. The following are members: Cystathionine gamma-lyase, Cystathionine gamma-synthase, Cystathionine beta-lyase, Methionine gamma-lyase. Aminotransferase class I and II;Interpro entry IPR004839. Molecular function - transferase activity, transferring nitrogenous groups (GO:0016769); biological process biosynthetic process (GO:0009058). 2 putative N-glycosylated sites at positions 76-173, 1 putative O-glycosylated site at position 175 is predicted.
CbRS-2	1175	EU878242	possible cell adhesion protein"	Conceptual protein by similarity; seems to be related to homologs of fasciclin-I, which are proteins involved in cell adhesion. Fasciclin domain. This extra cellular domain is found repeated four times in grasshopper fasciclin and as well as in proteins from mammals, sea urchins, plants, yeast and bacteria; pfam02469. This product is missing around 15 amino acids. intrinsic disorder 1 17 , Region name: "Fasciclin" FAS1 48 149 , FAS1 186-283 Fasciclin domain. The encoded product contains 2 putative N-glycosylated sites at positions 24 and 104.
CbRS-3 *	1106	EU878243	Predicted protein	No significant homology was found. Signal peptide 1 28, most likely cleavage site between pos. 18 and 19: ATA-DC, 1 putative TM helix at positions 13-35
CbRS-4	1036	EU878244	Predicted protein	
CbRS-5 *	920	EU878245	Predicted protein	No significant homology, signal peptide 1-35, transmembrane helix 41-60, transmembrane helix 72-94, transmembrane helix 109-131, one N-glycosylation site at position 120 and most likely cleavage site between pos. 35 and 36: VAA-SP. Contains 6 putative phosphorylation sites at positions 36, 40, 64, 134, 147 and 67. The predicted product has low index of antigenicity, highly hydrophobic. Calculated Molecular Weight = 18138.22 Estimated pI = 9.83
CbRS-6	759	EU878246	ATP Synthase, H+ transporting	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C2 isoform. Cellular component membrane (GO:0016020) Biological process ATP synthesis coupled proton transport (GO:0015986), Cellular component proton-transporting two-sector ATPase complex (GO:0016469) Molecular function hydrogen ion transporting ATPase activity, rotational mechanism (GO:0046961). Molecular function hydrogen ion transporting ATP synthase activity, rotational mechanism (GO:0046933)
CbRS-7*	756	EU878247	40S ribosomal protein S4	Structural constituent of ribosome. Involved in process such as protein biosynthesis. S4/Hsp/ tRNA synthetase RNA-binding domain; The domain surface is populated by conserved, charged residues that define a likely RNA-binding site; Found in stress proteins, ribosomal proteins and tRNA synthetase.
CbRS-8	731	EU878248	Eukaryotic translation elongation factor 2	Position 2 to 63 Elongation factor G, domain IV. "This family represents domain IV of archaeal an eukaryotic elongation factor 2 (aeEF-2) and of an evolutionarily conserved U5 snRNP-specific protein. U5snRNP is a GTP-binding factor closely related to the ribosomal translocase EF-2. In complex with... cd01681" /db_xref="CDD: 58276"7 65 to 154 EFG_C Elongation factor G C-terminus This domain includes the carboxyl terminal regions of Elongation factor G, elongation factor 2 and some tetracycline resistance proteins and adopts a ferredoxin-like fold.
CbRS-9 *	718	EU878249	Minicollagen 1	**Very interesting gene.** The predicted protein contains regions rich in histidines followed by at least 1 lysine. By similarity, nematocyst-specific, major component of the nematocyst wall. Within the *M. kingi* EST collection there is a full length clone (Mk-EST_512). In contrast with other genes (such as Cb-CSL AV-9- C3 –like gene or Cb-CSL AV-11-vitellogenin), this minicollagen is more similar to Hydrozoa than to Anthozoa. Homologs are found in *Hydra magnipapillata* (pir||A41132), *Acropora donei* (dbj|BAA06407.1|) and , *Clytia hemisphaerica* (gb|ABY71252.1|).
CbRS-10 *	644	EU878250	Ferritin	Full encoding region containing clone. No 5' UTR, initial methionine, poly A tailed. Protein involved in ferric iron binding, process- iron ion homeostasis. Protein contain classic eukaryotic Ferritin (Euk_Ferritin) domain. **Very interesting finding-** In a EST analysis of adult salivary glands of toxic ticks, ferritins were found. (Mans *et al.* 2008)
CbRS-11	568	EU878251	Ribosomal protein L16p/L10e	Incomplete cDNA seems to be missing 130 amino acids and 5'UTR. It is a homolog of the putative tumour suppressor QM in the marine sponge *Suberites domuncula.* Intracellular product. The function is structural constituent of ribosome, involved in protein biosynthesis process.
CbRS-12	1617	EU878252	Flap endonuclease	Full-length clone. No completely sequenced. 5' and 3' end sequence available. Observed 5' and 3' UTR and poly A tail. Eukaryotic enzyme that functions in nucleotide-excision repair and transcription-coupled repair of oxidative DNA damage. Contains a metal binding site. It has 82 and 74 % of similarity with flap endonuclease-1 *N. vectensis* and *Aedes aegypti*.

Table 2. 6 *M. kingi* cDNA clones immunoreactive towards CSL *C. fleckeri* antivenom (CSL AV)

Rows shaded in yellow correspond to either full-length cDNA clones or complete encoding regions. White rows correspond to partial cDNAs (lacking either 5' or 3' region). Gaps and cDNA inserts were estimated by digestion assay using either combination of restriction enzimes within the pBSK- polylinker. Red asterisk * = proteins that undertake a classic or non-classical secretion pathway were predicted using SignalP and SecretomP 2 obtaining neural network values exceeding the normal threshold of 0.5 (Bendtsen *et al.* 2004). Conceptual translations were searched in diverse data bases: Simple Modular Architecture Tool (SMART) (Schultz *et al.* 1998), Protein Family data base (Pfam) and Interpro entry were used to enrich annotations. CnidaBase . Putative O-glycosylated and N-glycosylated sites were predicted using NetPhos 2.O and NetOGlyc 3.1 Server (Blom *et al.* 1999; Julenius *et al.* 2005). Rows shaded in orange correspond to partially sequenced cDNAs but the clone encodes a complete product (Predicted by gene similarity). The accession numbers correspond to the nucleotide sequence the protein accession number is pending.

Clone ID	bp	Accession	Final Annotation	Relevant details about the cDNA clone and the encoded product.
Mk-CSL AV-1	2397	EU878255	Actin	Full-length clone, completely sequenced. Long 5'UTR (1090 bp), Kozak (TACTCTTTCAACATGGCCGA), termination codon followed by 3' UTR and poly-adenylated clone. Actin is a ubiquitous globular protein, roughly 24 kDa. Actins are involved in the formation of filaments that are a major component of the cytoskeleton. Interaction with myosin provides the basis of muscular contraction and many aspects of cell motility. Actins are one of the most highly conserved proteins throughout evolution that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells. VERY ANTIGENIC
Mk-CSL AV-2	1639	EU878256	Elongation factor 1 alpha	Full length clone. cDNA containing 5 UTR, Kozak (ACTTTACCGAAATGGGTAAGG) sequence, termination codon and 3'UTR, poly-A tailed. It encodes a 468 amino acid residue product. Eukaryotic elongation factor 1 (EF1) alpha subfamily. EF1 is responsible for the GTP-dependent binding of aminoacyl-tRNAs to the ribosomes. This protein displayed a high antigenicity as determined using three methods; Parker = Parker antigenicity, Prot Ag = Protrusion Index antigenicity, Welling = Welling antigenicity. Involved in translation. VERY ANTIGENIC
Mk-CSL AV-3 *	1624	EU878257	TPA: gag-pol polyprotein	Not fully sequenced cDNA of at least 2000 bp;. poly-adenylated. Most likely to be a Reverse transcriptase (RTs) in retrotransposons. Members of this family represent the RT domain of a multifunctional enzyme. Gap
Mk-CSL AV-4	1629	EU878258	Similar to non-muscle myosin heavy chain	Not fully sequenced. 5'/3' UTR and polyA tail observed. It has been directionally joined at nucleotide 785; the gap is of unknown length, several internal sites similar to the pBSK- polylinker are observed. Digestion assays suggest a cDNA clone of 2000bp. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains Myosins are fundamental contractile proteins highly conserved in all eukaryote cell types. VERY ANTIGENIC. GAP
Mk-CSL AV-5 *	686	EU878259	Cyclophilin A	**Very interesting gene.** cDNA fully sequenced. No 5' UTR, mid-strength Kozak Sequence: (AACAAGAAAATGG), complete coding region, 3' UTR and poly-adenylated. Highly similar to peptidyl-prolyl cis-trans isomerase, mitochondrial precursor (PPIase) (Rotamase) (Cyclophilin F) and to cyclophilins are involved in **immune response.** VERY ANTIGENIC.
Mk-CSL AV-6	563	EU878260	Ribosomal protein L21	Full length cDNA clone comprising both UTRs, complete coding region and polyA tailed. Encodes a product of 161 amino acid residues. The calculated molecular weight is 18416.03 and an estimated pI = 10.95. Valine, glycine, lysine and arginine are more represented than the rest of the residues. Ribosomal proteins L21 are fairly well conserved proteins with an average similarity. The gene ontology component for homologs proteins is intracellular; ribosome; gene ontology functional category is structural constituent of ribosome and is involve in protein biosynthesis processes. This protein is clearly phylogenetically related to the *N. vectensis*. VERY ANTIGENIC
Mk-CSL AV-7	453	EU878261	Ribosomal protein L35Ae	Fully sequenced and cDNA clone, containing 23 bp upstream the ATG codon. Kozak sequence (CTAAAGCAAAAGGATGGACCAGA). Complete encoding products homologs to this cDNA are found in *N. vectensis* and *Hydra vulgaris*. Involved in translation. VERY ANTIGENIC
Mk-CSL AV-8	1307	EU878262	similar to HSP90-like ATPase. TNF receptor-associated	Truncated cDNA clone, fully sequenced. 5' UTR and Kozak sequence were not observed. Termination codon and poly A tailed. The ORF encodes a product of 399 amino acids. Phylogenetic comparisons indicated lack of approximately 320 amino acids upstream. Belongs to the heat shock protein 90 Superfamily. (HSP90), Molecular chaperones of 704 residues. Regulatory functions. ANTIGENIC
Mk-CSL AV-9 *	943	EU878263	Predicted protein	**Very interesting gene.** Truncated clone. 750 bp of coding region followed by 3' UTR and poly (27) A tailed. The first 100 amino acids comprise a transmembrane region followed by repetitions containing arginine, glycine, isoleucine and phenylalanine (38.96, 11.65, 8.3 and 11.65% of amino acid composition, respectively). Interestingly the message seems to have an additional translational start site, encoding a -2 frame protein with RLPNTLRLPKTLRLPNTLRLPNILRLMKTLRLP repeats; considerably high antigenicity is found flanking the repetitions. (Park antigenicity, 3.5 and protrusion index 5 for the downstream region). In the -2 frame the most likely cleavage site between residues. 32 and 33: ANS-EK. Found similarities with several parasitic animals, such as Plasmodium. The sequence also was found to be similar to *Candida* **agglutinin-like protein** (ALS), This proteins **mediate humoral immunity in invertebrates.** This family consists of several agglutinin-like proteins from different *Candida* species. ALS genes of *Candida albicans* encode a family of cell-surface glycoproteins with a three-domain structure. MAY BE TRULY ANTIGENIC
Mk-CSL AV-10	672	EU878254	Hypothetical protein	Truncated cDNA clone, fully sequenced, no 5' UTR, no iM, encoding region lacking 360 to 370 amino acids, 3' UTR and poly-adenylated. Encoding region is similar to a transmembrane protein, found in platypus liver, lung, brain and spleen tissues. No known function. Predicted to be involved in energy metabolism and immune response.
Mk-CSL AV-11 *	647		Trypsin-like serine protease	A truncated, inverted and not fully sequenced cDNA clone, poly A (18) tailed. The trypsin-like serine protease (Engel *et al.*) domain is not complete. Many TSPs are synthesized as inactive precursor zymogens that are cleaved during limited proteolysis. Trypsin activity involved in proteolysis and stress response involved.

2.4.9 Lack of molecular data in Cubozoa: The allergens and chaperones dilemma

The cnidarian repertoire of molecular structures able to elicit allergic reactions is unknown except for an allergen-causing protein reported in the red soft coral *Dendronephthya nipponica*. This allergen, Den n1, is able to induce a strong and immediate intradermal reaction (Onizuka *et al.* 2001)**.** No allergen has been identified in members of the subphylum Medusae. Moreover, although six venom proteins from cubozoans have been characterized, all of them are clustered into one family (cytolytic box jellyfish toxins). Other than that there is no molecular data relating to cubozoan allomones available to allow comparisons. This makes it more difficult to draw useful and valid inferences and only homolog comparisons within cnidarians and with other organisms can give an insight into cubozoan toxins. Furthermore, due to the ancestral position of cnidarians in the evolutionary scale, many genes found in cubozoans have homology with predicted and hypothetical proteins from *N. vectensis* and hydrozoans but lack homology with bilateral animals, so that molecular characterization of those sequences is hampered.

Nevertheless, the finding that several antigens (e.g. several ribosomal proteins, cyclophilins, pathogenesis-related plant proteins, actin binding proteins) from fungi, plants and animals that cause human allergies have been identified by a similar approach (Achatz *et al.* 1995; Bachrach *et al.* 1997; Cadot *et al.* 2000; Flueckiger *et al.* 2002; Mittermann *et al.* 2005; Saxena *et al.* 2003) and the fact that ribosomal proteins (Mk-CSL AV-6, Mk-CSL AV-7 (L21) and a member of the cyclophilin family (Mk-CSL AV-5) were found in this research might represent a possibility of genuine cross-reactivity. It would be interesting to produce one of these putative allergens *in vitro* and use it in functional assays to determine human reactions.

Other incomplete clones related to ribosomal proteins such as Cb-BJA-2 (RP L10/QM like) were also seen. Therefore RP fusion protein expression was dismissed and the only explanation for this clone to be antigenic was for it to react with bacteria, phage or mistakenly identified as immunoreactive due to a membrane background.

CSL AV is a complex antibody resource, likely to be produced with impure venom preparations (tentacle and mucus contamination) and degraded proteins. It has been suggested it lacks medically important factors (Endean & Sizemore 1988). The percentage of specific nematocyst protein antibodies present in CSL AV may not represent the complete set of venom proteins as a result of low venom protein antigenicity therefore eliciting weak IgE responses in immunized sheep leading to low concentrations of those antigen-specific antibodies in CSL AV. This assumption is supported by western blot analyses of native venom proteins in which only few venom proteins were antigenic against CSL AV. In addition, components other than nematocyst proteins (*e.g.* constituents with mucus origin) may act as immunodominant antigens. This is of importance because mucus plays an important role in immunity as it represents the first physicochemical barrier to invasion. Mucus may contain molecules with an action mechanism involved in depressing the aggressor or prey immune system such as virulence factors and allergens.

Sheep sera components may be another factor that led this experiment to be ambiguous and difficult to interpret, and perhaps it represents the major weakness in the immunoscreening approach taken. The heterogeneity of CSL AV would limit the selection of specific phagemids due to excessive binding to unwanted gene products (as seen for the large number of false positive clones isolated and characterized). On the other hand, other mucus components of importance in phagogenicity and allergenicity could elicit immune responses in immunized sheep and this possibility should not be underestimated and further research into cubozoan allergens is encouraged.

2.4.10 Polyclonal antibody probes others than CSL AV

From the fourteen *C. barnesi* reactive clones detected with *C. barnesi* polyclonal antibodies of whole jellyfish extract raised in rabbit (Table 2.5), two clones were of interest: a full length ferritin homologous to that found in an EST analysis of a salivary gland of ticks toxic to humans and CbRS-9, which encodes a minicollagen that it is involved in nematogenesis (production of the stinging cells). Minicollagens are also the major constituent of the walls of nematocysts in hydrozoans (Denker *et al.* 2008; Engel *et al.* 2002; Engel *et al.* 2001) and anthozoans (Wang *et al.* 1995). Both ferritins and minicollagens are expected to be expressed abundantly, therefore increasing the chance of being selected as false positives. Although CbRS-9 could be a false positive, it is in fact a relevant clone and it may have similar signal peptides and propeptide regions to those found in the proteins inside the nematocyst capsules, and thus have some similar antigenic epitopes. However CbRS-9 is a truncated cDNA and consequently no antigen could be expressed. Therefore the antigenicity observed is likely due to cross reactivity with phage.

In the search for a more specific antibody probe, screening of Irukandji cDNA expression libraries using *C. barnesi* and *M. kingi* species specific polyclonal antibodies towards Irukandji nematocyst proteins seemed to be particularly attractive. But when tested in immunoscreening plates no reactive phage plaques were identified. Similar results were obtained with human antibodies in spite of both probes binding to proteins present in crude jellyfish extracts. Species-specific antibodies (raised either in mice or rabbit) and human sera are complex mixtures of antibodies that cross react with proteins present in jellyfish extracts. However, the levels of reactivity were not strong and importantly, not specific enough to detect recombinant proteins expressed in the cDNA expression library. These results indicate that polyclonal probes others than CSL AV are not suitable to screen jellyfish expression libraries.

2.4.11 Key parameters to take into account to interpret outcomes.

Ideally immunoscreening of cDNA libraries is a highly sensitive method and has the potential to allow rapid isolation of antigens and allergens. However, the success of this approach is highly dependant on the specificity of the antibodies. This is

determined by the capacity of jellyfish proteins to act as allergens and induce a strong immune response. Other factors such as tissue optimized libraries and high quality of starting material, must be taken into account. The technique is particularly useful for antibodies that are strongly reactive by immunoblotting (Figure 2.1 and 2.2), but importantly it is limited to antibodies that are reactive against linear epitopes. This is of relevance and may have influenced the outcome of these experiments. Non-denatured proteins were used in the production of CSL AV and the polyclonal antibodies used, whereas recombinant proteins expressed through expression libraries are generally linear (Knudsen & Vedeler 2006) which suggests that epitope recognition did not take place.

The lack of jellyfish samples and the fact that *C. barnesi* and *M. kingi* possess nematocysts in both the umbrella and the tentacles, led to the use of RNA from whole jellyfish samples for the construction of cDNA libraries. This means that perhaps a library specific for cnidoblast cell containing tissue should be considered in further research thus narrowing the search for epitopes of interest. Nevertheless both of the Irukandji libraries have good clonal representation and assuming that transcripts encoding venom proteins are not rare, it is likely that a sufficient number of cDNA clones encoding functional venom-related proteins occur in both of the phage display libraries. It may be that expression is largely inhibited because the fusion proteins may be detrimental to the host cell growth or other unknown factors.

During immunoscreening experiments, some cross reactivity with bacterial and/or phage proteins was observed (data not shown in tables). This could be avoided in future research by treating the antibody solution to filters exposed to bacterial lysates prior to use in immunoscreening experiments. In the case of human sera a more specific probe could be obtained by purifying the IgG and IgE.

In relation to the antigenic reactions detected in extracts from cubozoan nematocysts towards polyclonal antibodies, it is speculated that with the exception of a few antigenic and immuno-dominant proteins in the range of 42 to 46 kDa and some other minor bands of 120 and 20-30 kDa (See lines 7 and 10 of Figure. 2.1), most venom proteins are likely to exhibit poor cross reactivity. Although some of the nematocyst

proteins could lack or have poor antigenic epitopes, this should not be misinterpreted, because cubozoan toxins may elicit a strong immunological response in envenomated humans that may be independent of antigenic epitopes but may otherwise be molecules acting as pathogens or mimicking molecules involved in immune cell communication. For example, inducing uncontrolled production of cytokines that activate T cells and macrophages producing what is known as “a storm of cytokines”. The symptoms of a cytokine storm primarily are fever, swelling, redness, fatigue, nausea and gastrointestinal discomfort similar to those observed in cnidarian envenomation and in severe cases, a cytokine storm can lead to systemic dysfunction of organs. In addition it is likely that medically important but not abundant constituents of cubozoan venom, such as peptides targeting excitable membranes (e.g. cysteine-rich peptides similar to the *Stichodactyla helianthus* neurotoxin and its homologs in other cnidarians) were missing in the probes used, possibly as a result of poor capability to induce antibodies and also because the representation in the whole venom may inherently be low. Furthermore, there may be some other proteins that, although abundant in the venom, lack antigenic epitopes or possess epitopes that are not well exposed, consequently exhibiting low antibody production levels in immunized organism.

2.5 CONCLUSIONS

The antibody strategy undertaken in this investigation has several limitations inherent to the lambda ZAP cDNA expression library system itself, as well as being restricted by the use of *E. coli* as an expression host. This is primarily the potential toxicity of gene products to bacteria and the limited folding capacity in an *E. coli* expression system. Another limitation was the use of non specific antibodies and the lack of bacteria and phage antibody pre absorption procedures before screening.

The attempts made to obtain cDNAs encoding proteins involved in defence mechanisms are not easy to interpret as a consequence of there being many non specific antigenic cDNA clones. Whether or not some of the full length clones, specifically the clones encoding ribosomal proteins and cyclophilin, that resemble

those found to act as allergens in other organisms play a role in jellyfish envenomations is unknown.

It is suggested that complex antibody probes are not suitable for detecting Irukandji toxic components expressed via cDNA expression libraries. It is in fact likely that some of the jellyfish toxic proteins are detrimental to the host cells. The main contribution of this research chapter was to create awareness of the poorly antigenic nature of jellyfish venom proteins, with the exception of the immunodominant 42-46 Cubozoan cytolytic protein family (shared in five cubozoans) and a few other antigenic proteins. We must be aware of this possibility if pursuing research to improve the antigenicity of a potential antivenom for use in the treatment of envenomations by cubozoan venom proteins.

Finally, in spite of the uncertain outcomes of this study, the results reported from these experiments produced a significant amount of transcriptome information from two medically important Irukandji jellyfish.

CHAPTER 3

Expressed Sequence Tag survey and comparative gene analysis from a cubozoan responsible for Irukandji syndrome in Humans.

3.1 ABSTRACT

The Australian box jellyfish *Malo kingi* is a rare Irukandji stinger, likely to be responsible for a fatality in North Queensland in 2002 (Gershwin 2007b). Despite its medical importance, surprisingly nothing is known about its toxin composition. We attempted to address this issue by generating a survey of 1,644 expressed sequence tags (ESTs) from a non-normalized, unidirectional cDNA library. After processing the raw data, a total of 1,476 high quality tags were assembled resulting in 1,039 clusters of which 818 were singletons. Approximately 5.38% were clusters related to reverse transcriptase-like enzymes, endonucleases, reverse transcriptases or retrotransposons. Fifteen per cent of the deduced peptides are predicted to be expressed via a secreted pathway, 50% of the conceptual encoded products do not have homologs in current databases. According to comparative gene analysis, the cubozoan data set showed higher similarity with anthozoan rather than to hydrozoan genes, and so has greater similarity with vertebrates rather than with invertebrates such as flies, nematodes or comb jellies (ctenophores), previously reported for other cnidarian species (Kortschak *et al.* 2003). A small portion of the data revealed true similarity with non-metazoan genes (e.g. bacteria, fungi, plants and viruses); the presence of such genes has been previously found in anemones, corals (Technau 2005) and in hydra.

The generation of ESTs allowed us to identify several genes that are likely to be associated with defence, stress mechanisms or immune-function and genes involved in developmental pathways. Among the most relevant genes are two cytolysins (EST_269 MkTX1 and 2), a metalloprotease (EST_136), a serpin (EST_540), a metalloserpin (EST_617), Mk-TSP-2 (EST_63/1162), a cystatin (EST_458), a pathogen-related protein (EST_310), a dermatopondin (EST_906) and three protein

members of the ShK toxic family (EST_332-1/2 and EST_93). Interestingly, the last six genes encode products containing a six cysteine motif that was originally described in an ShK toxin protein from the sea anemone *Stichodactyla helianthus* The ShK toxin domain appears as a single, double or a tandem arrangement of three to four motifs. In addition, we detected other important genes such as VDAC, cyclophilin A, caspase 8 and transcription factors.

EST approaches are effective for the identification of defence genes in cubozoans and serve to reveal important information on the biology of cubozoans. Although a small collection, this study provides the first set of genetic data for a rare and likely lethal box jellyfish from Australian waters and it is of significant value for further research.

***Keywords**:* Irukandji; box jellyfish; Cubozoa; expressed sequence tags; defence mechanisms. ShK domain containing proteins.

3.2 BACKGROUND

Cnidarians are among the earliest living metazoans. It is generally accepted that the phylum split from the stem of the evolutionary tree that gave rise to protostomes and deuterostomes around 800-1,000 million years ago (Morris 1993). Early metazoan lineages are of key importance for unravelling what could be the metazoan predecessor gene pool. Over the last decade, rapid convergent information initially based on ESTs from cnidarian research groups has provided insight into the origins of ancestral molecules. These demonstrate that important molecular mechanisms implicated in cellular signalling cascades as well as transcription factors that are involved in early patterning and development are found in earlier phyla than had been classically assumed. Some examples are genes involved in development such as *Hox*, *ParaHox*, *cnox* and *Wnt* genes (Cartwright *et al.* 2006; Finnerty & Martindale 1999; Galliot & Miller 2000), axial patterning (de Jong *et al.* 2006), neurogenesis, regeneration (Guder *et al.* 2006; Holstein *et al.* 2003; Miljkovic-Licina *et al.* 2004), and genes involved in vision such as *Pax* genes (Garm *et al.* 2007b; Kozmik *et al.* 2008; Nordstrom 2003). In addition to these, gene products that form ions channels (Jeziorski *et al.* 1998) and gene products involved in the innate immune system in lower metazoans (Dishaw *et al.* 2005; Hemmrich *et al.* 2007; Miller *et al.* 2007). The outcomes are of major significance not only because they challenge long held views on animal evolution (Darling *et al.* 2005), but because they show a remarkably high level of genetic complexity in cnidarians (Miller *et al.* 2005; Technau *et al.* 2005).

It is important to mention that all cnidarians, with the exception of Class Anthozoa (anemones, soft and hard corals) and a few hydrozoans, are united by the pelagic medusa stage and contain linear and often fragmented mitochondrial genomes (Bridge *et al.* 1992; Miller *et al.* 2005; Technau 2005). In contrast, Anthozoa have mtDNA in the form of circular molecules, which is similar to ctenophore and all other known metazoans, and for this reason it is suggested that Anthozoa is the basal class, while morphological innovations place Cubozoa as a modern class within the phylum. Other relevant contributions of genomic data from cnidarians is the finding of gene loss in

model animals such as *Drosophila melanogaster*, *Caenorhabditis elegans* or *Saccharomyces* (Kortschak *et al.* 2003) implying a high rate of divergence among lower metazoans. Another interesting feature is that cnidarians have maintained non metazoan genes, for example genes from fungi, prokaryotes, plants and protists, in their genomes (Technau *et al.* 2005). On the other hand a high proportion of genes implicated in human disease are present in the genome of basal cnidarians, including some that are entirely absent in well established animal models. Cnidarians, therefore, are able to provide important insights into the mechanism and treatment of human diseases (Sullivan & Finnerty 2007).

ESTs are randomly selected 5′ or 3′ end, single pass sequences of short length (100-800 bp), from cDNA libraries of specific tissue or whole organisms. EST projects are usually a time and cost effective strategy for gene discovery. This was indeed the reason that lead ESTs to initially be used as a primary resource for human gene discovery (Adams *et al.* 1991). Presently, ESTs and cDNA sequences remain an important resource for rapid transcriptome exploration. This genomic approach has been also widely used to rapidly identify and characterize defence mechanism encoding genes from venomous animals involved in human envenomation such as ticks (Batista *et al.* 2008), snakes (Wagstaff & Harrison 2006), cone snails (Pi *et al.* 2006), spiders (Liping J *et al.* 2008), venomous fish, bees (Whitfield *et al.* 2002), dinoflagellates (Leggat *et al.* 2007), bacteria and fungi.

In recent years, genomic libraries and ESTs of cDNA libraries have proven to be a powerful aid for gene transcript identification in cnidarians. There are large raw data sets of genomic sequences and ESTs for *Hydra magnipapillata*, *N. vectensis*, *Acropora millepora*, *Acropora palmate*, *Porites lobata* (Miller *et al.* 2005), and a small collection of 2153 ESTs for *C. capillata* from a tentacle cDNA library (Yang *et al.* 2003a). Most of the molecular data of cnidarians is available through private collections but mainly from public databases such as CnidBase, StellaBase, Hydra EST database and the dbEST NCBI. The full genomic sequence of the sea anemone *N. vectensis*, a basal representative of the Cnidaria phylum, has also recently been published (Putnam *et al.* 2007).

Among the genomic information for cnidarians, it is significant that there are no more than 25 mRNA sequences at public databases derived from Cubozoa. Those sequences that have been identified and characterized include: *C. rastonii* CrTXs, *C. alata* CaTX-A (Nagai *et al.* 2000a; Nagai *et al.* 2000b), *C. quadrigatus* CqTX-A (Nagai 2003) and two cytolytic and major proteins contained in *C. fleckeri* venom (Brinkman & Burnell 2007). Due to the remarkable visual system of Cubozoa, with 24 eyes distributed in four sensory structures –rhopalia- (Garm *et al.* 2007a), *Tripedalia cystophora* has been a model Cubozoan for molecular, evolutionary and developmental studies characterizing the production of vision in lower metazoans. Therefore mRNA sequences of crystallin genes, *Pax* transcription factors and photoreceptors have been obtained and used for analyzing evolutionary relationships.

As time passes, the amount of molecular data increases enormously for cnidarian species but not for Class Cubozoa. The main aim in this chapter was to focus on generating good quality molecular data on the proteins of Irukandji jellyfish, with emphasis on genes coding for proteins involved in defence. The molecular identification of these genes would lead to *in vitro* over expression of proteins that may give insight into the overall process of Cnidaria envenomation and perhaps in the future, antivenom technologies could benefit from genomic approaches to identify and characterize target genes for use for potential therapeutics.

3.3 EXPERIMENTAL PROCEDURES

3.3.1 RNA isolation, cDNA library construction, mass excision and DNA sequencing.

Two whole adult specimens were disrupted with glass pearls in Trizol (Invitrogen) and total RNA was isolated as described by the manufacturer. mRNA was isolated using PolyA Track Kit (Promega). The cDNA library was constructed using a ZAP-cDNA Synthesis Kit and Gigapack® III Gold Cloning Kit (Stratagene; La Jolla, CA) following the manufacturer's specifications. The phage library fractions were transformed into a phagemid library using the mass excision protocol. Transformed *E.*

coli were plated on agar with 60µg.ml^{-1} carbenicillin selection. Agarose plates were incubated overnight at 37°C, and randomly chosen colonies were used to inoculate cultures grown in standard conditions. Plasmid DNA was purified (QIAGEN), insert lengths were estimated by digestion with restriction enzymes within the pBSK-polylinker (Stratagene) and resolved by 1% (w/v) agarose gel electrophoresis. Gels were stained with ethidium bromide and visualized under UV light.

Clone redundancy was tested by randomly choosing two plasmids (1 and 0.5 kb), labelling them with ^{32}P, and using the probes to screen the first 100 plasmid DNA samples. Generally, cDNA clones were sequenced using the T3 universal primer. Some of the clones that had good quality sequences and were relevant were selected for reverse sequencing with the T7 universal primer.

3.3.2 ESTs cluster assembly

Raw sequence-trace data files were processed using Sequencher (GeneCodes Corp.) in order to remove low quality sequences as well as vector contamination. Any sequences that were less than 150 bp were eliminated from subsequent analysis. Contaminating regions of *E. coli* and bacteriophage λ DNA were detected with the algorithm BLASTN from the NCBI web site http://www.ncbi.nlm.nih.gov. The remaining sequences were used for cluster assembly with SeqMan (Lasergene software, DNASTAR), using default Pro Assembly Parameters.

3.3.3 Annotation and functional classification

Comparative analysis using the NCBI database commenced on the 10th of October and finished on the 26th October, 2008. Consensus sequences were exported to a FASTA format document. The ten best hits for BLASTN, BLASTX and BLASTP were obtained by remotely running Blastcl3 at NCBI, using default parameters. The predictions of translation reading frames were achieved by submitting the EST data to an ORF predictor interface hosted at https://fungalgenome.concordia.ca, using a default threshold of 1 x 10^{-5} as a guide. Signal peptide recognition sites, non-classical and leaderless protein secretion were predicted with SignalP 3.0 and SecretomeP 2.0

servers respectively (http://www.cbs.dtu.dk/services/SignalP/ and http://www.cbs.dtu.dk/services/SecretomeP/) (Bendtsen *et al.* 2004a). Blast2GO (http://www.blast2go.de/) and ProtFun 2.2 Server (http://www.cbs.dtu.dk/services/ProtFun/) were interrogated for prediction of function, enzyme class and gene ontology category.

3.4 RESULTS AND DISCUSSION

3.4.1 M. kingi *cDNA library*

A non-normalized and amplified cDNA library with a titre of 9 x 10^{10} *pfu*/ml was constructed for *M. kingi*. The frequency of vector containing inserts was 90%, with a cDNA length distribution range between 0.5 and 4.5 kb and an average length (n = 200) of 1,727 bp. Clonal representation in the library was homogeneous as seen in the cluster distribution in Figure 3.1.

3.4.2 Sequencing, assembly of ESTs and cluster identification by homology analyses.

To reduce data redundancy, improve sequence quality, base accuracy and transcript length and determine gene representation, ESTs were grouped by sequence identity. The original data set comprised 1,644 tags from which 125 sequences were either less than 250 bp or of poor quality. Forty-three sequences were identified to contain M13 phagemid or vector and were not used in further analysis. Finally 1,476 good quality sequences were selected to be clustered using Seqman, resulting in 1,039 clusters, from which 818 clusters were singlets (*i.e.* contained only one EST).

The overall prevalence of transcripts was homogeneously representative. The cluster size reflected the relative mRNA population since the *M. kingi* cDNA library was non-normalized and, because it was created with a whole animal rather than venom containing tissue such as tentacles or glands, a more diverse data set was obtained. The ten most abundant transcripts (with exception of cluster 206) were structural or

housekeeping genes, all previously characterized and found in all metazoans, such as actins, tubulins and myosins (see Table 3.1).

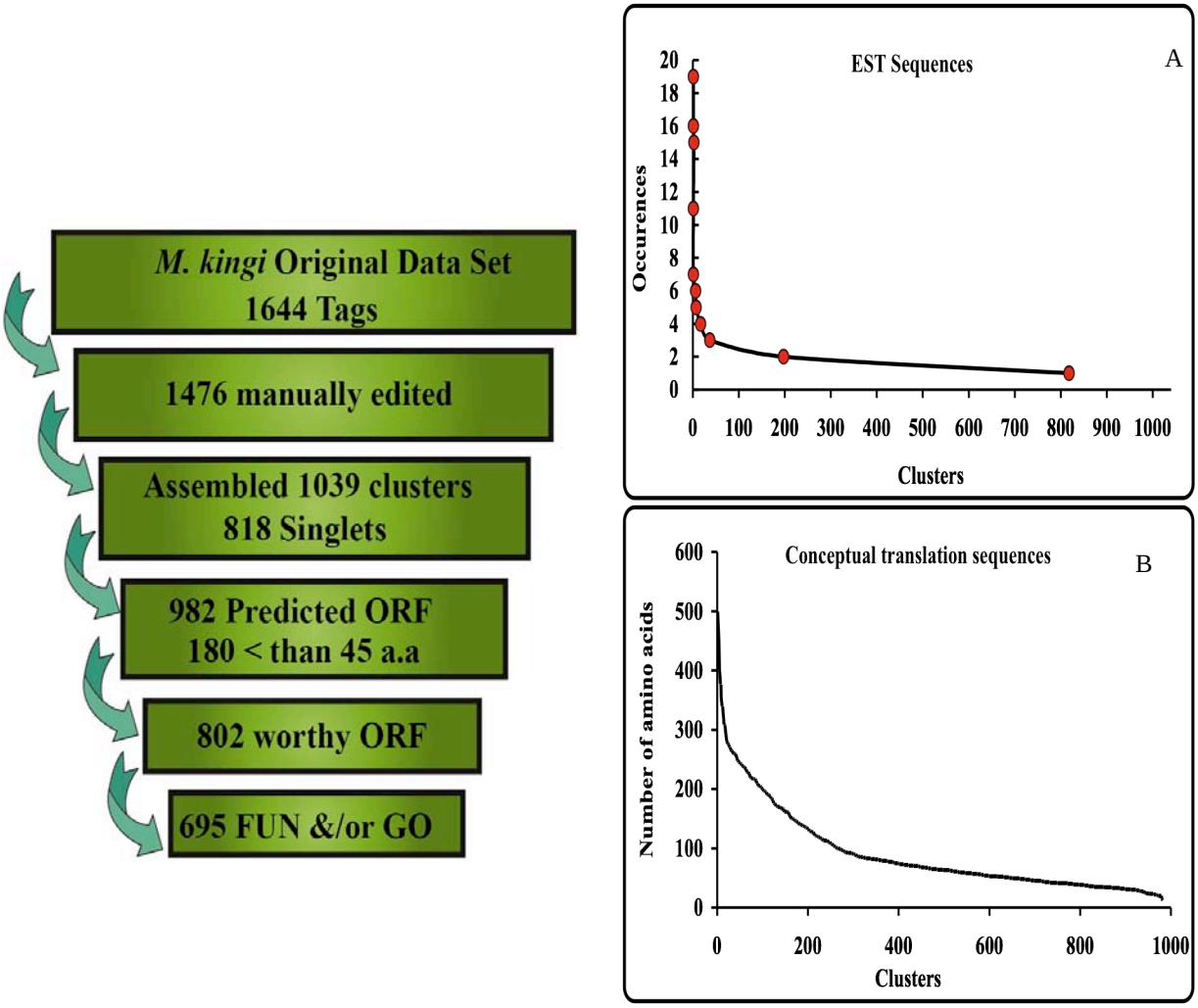


Figure 3.1 *M. kingi* flow diagram for EST processing, occurrence and distribution graphs

The flow diagram at the left indicates the strategy followed to process the *M. kingi* raw data to obtain gene function and gene ontology predictions. In A) EST cluster size distribution, representing the occurrence of tags within each cluster. B) distribution of sequences according to the deduced amino acid sequence length.

Table 3.1 Characteristics of the ten most abundant *M. kingi* transcripts

The length of the transcript is indicated in base pairs (bp), and the deduced amino acid sequence length by (a.a). Protein sequences containing complete open reading frames are indicated by (FL, full length). The ten clusters represent ubiquitous structural and hypothetical proteins. Transcripts that were also among the ten most abundant transcripts in the tentacle EST collection of the scyphozoan, *C. capillata* (Yang *et al.* 2003a), are shaded in green.

Cluster ID	*bp*	*a.a*	*No. Tas*	*Best hit BLASTP*	*Organism*	*Accession*	*e-Value*
215	3692	376(FL)	19	Actin, non-muscle 6.2	*Galaxea fascicularis*	BAC44866.1|	0.0
69	733	176 (FL)	16	Ferritin	*Dermacentor albipictus*	AAQ54711.1|	0.0
19	1572	450 (FL)	15	Alpha-tubulin	*Aurelia aurita*	AAP49551.1|	0.0
1038	1646	468 (FL)	11	Elongation factor 1 alpha 2	*Axinella verrucosa*	CAE45763.1	0.0
562	1128	326	7	Orf1ab polyprotein			0.15
138	1001	171	6	No hits found			
214	675		6	Hypothetical proteins	*Ornithorhynchus anatinus*	XM_001511968.1	1e-21
210	1380	278 (FL)	6	40S ribosomal protein S2	*Tribolium castaneum*	XP_975415.1|	e-110
206	802		6	No hits found			
208	1271	134 (FL)	6	ATP synthase		XP_001137325.1	7e-30

It was found that the cubozoan data set displayed more matches with higher organisms, such as human and other vertebrates, than to invertebrates, consistent with previous findings in other cnidarians, such as in the staghorn coral *Acropora millepora* and the starlet anemone, *N. vectensis* (Kortschak *et al.* 2003) and *Hydra vulgaris* (see Figure 3.2). In addition, during the time of analyzing the cubozoa EST library, the complete genome of several model invertebrates was of public acces at genebank, and during the last two years the genome of other invertebrates animal models were released *i.e* wasps, fruit flies and nematodes. The invertebrate data daset although smaller that the vertebrate countapart is fairly representated allowing fair comparision. Moreover, the analyses presented here was conducted masking repeated genes, thus by reducing redundancy more of the new data from invertebrates had better chance to appear as match vs the cubozoa data. Nevertheless, a bigger cubozoa data set could be a better indicator of gene lost trough the evolutionary history of some invertebrates phyla versus the phylum Cnidaria.

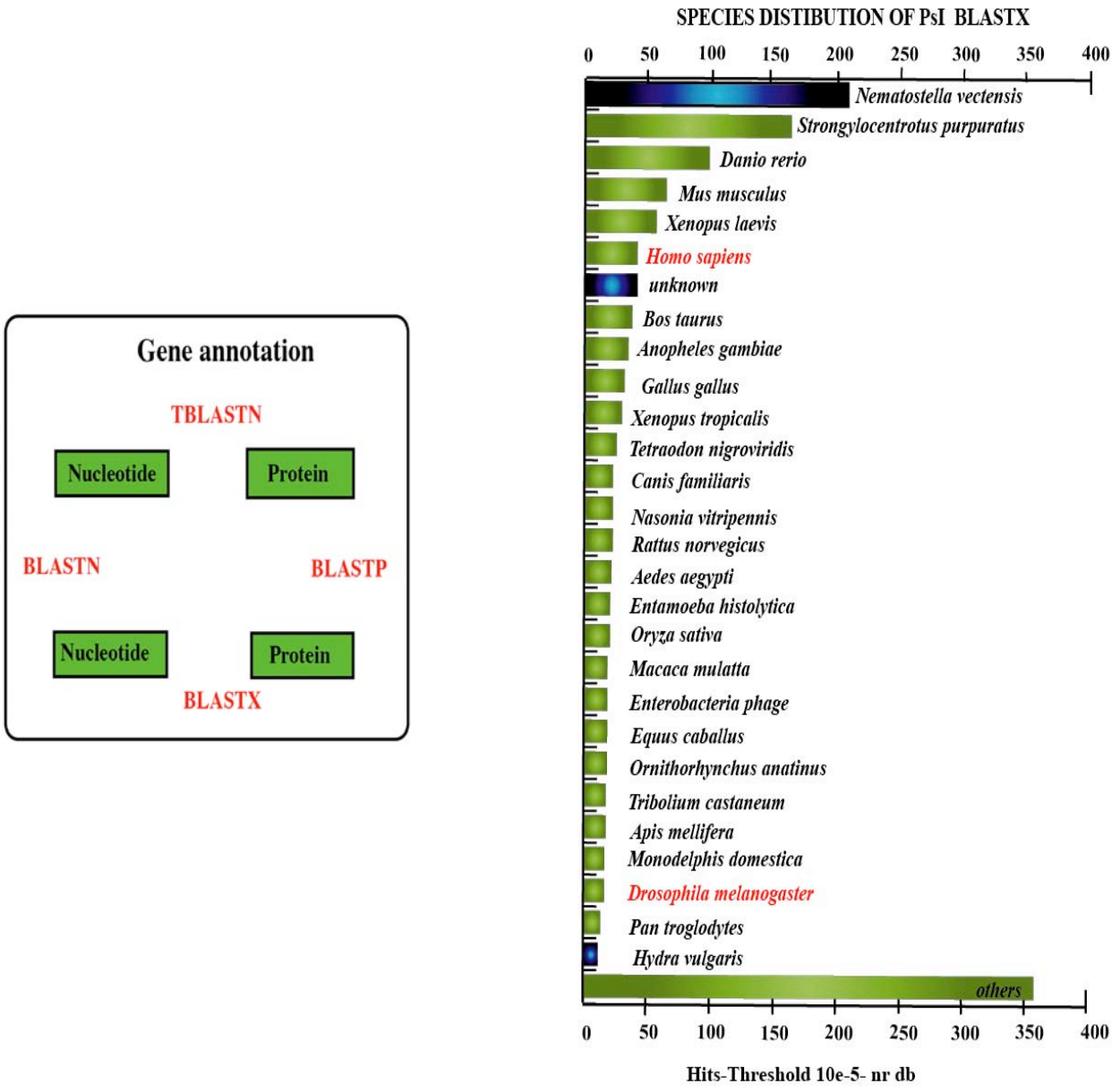


Figure 3.2 Gene annotation strategy and ESTs species distribution

Sequences were screened using BLASTX to identify contaminating DNA and BLASTP used to detect motif and protein domains. The species distribution graph of similar sequences was constructed based on BLASTX analyses using Blas2go. PsI stands for *M. kingi* before the species was described.

Unexpectedly, phylogenetic analyses by BLASTX showed that the cubozoan data set was more similar to the anthozoan *N. vectensis* than to hydra (Figure 3.2). It is likely to be an accurate representation due to the fact that this analysis was carried out prior to the release of the complete genome of *N. vectensis* and there was considerable

molecular data on hydra to allow fair representation for comparison. For each cluster, the ten best hits were retrieved from NCBI and all the data was manually inspected. It was observed that in most cases there was more similarity with the starlet anemone than with hydrozoan sequences; this result seemed to be trivial but in fact is of great relevance because for a long time the phylogenetic relationships among scyphozoans, cubozoans and hydrozoans were not strongly resolved and the consensus hypothesis was to separate hydrozoans from the other medusae. Werner (1973) championed the idea that cubozoans are more closely related to hydrozoans than scyphozoans. Although this view has been controversial, several authors favoured this affinity. However in a recent article there is reasonably strong evidence to place Stauromedusae, comprising stalked-sessile jellyfish as the earliest diverging lineage within Medusozoa, see figure 3.3 (Collins *et al.* 2006). The idea of the stalked jellyfish being the sister group is in fact interconnected with our results showing cubozoan genes are more similar to the basal anthozoan class rather than to hydrozoans which are in fact distantly related. (For further information, consult Collins *et al.* 2006; it contains a recapitulation and importantly a clarification of medusozoan phylogeny).

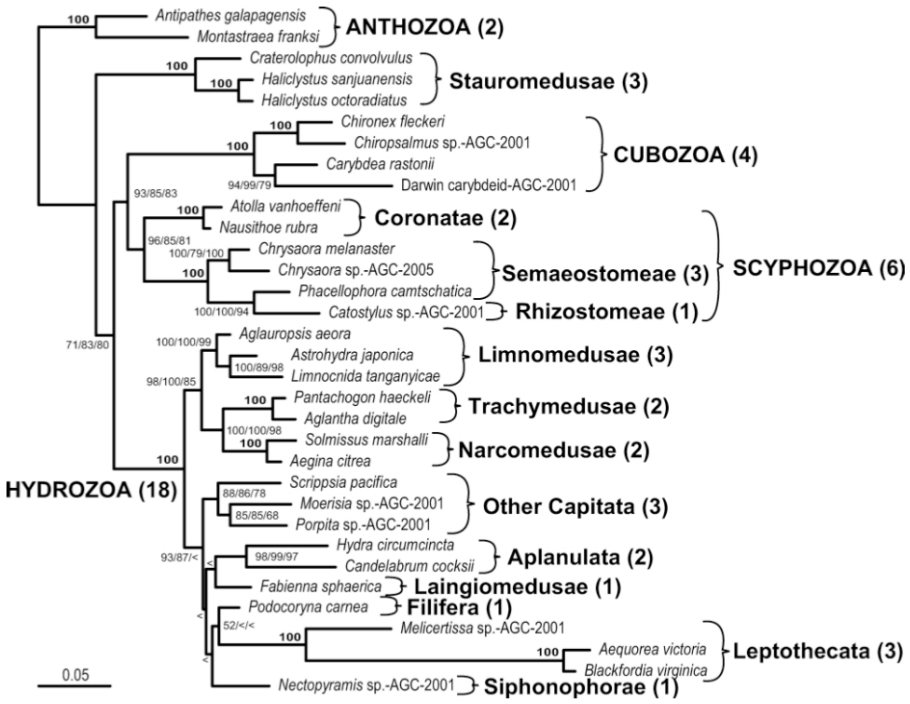


Figure 3.3 Phylogenetic position of Cubozoa within Phylum Cnidaria

Unconstrained topography has been taken entirely from (Collins *et al.* 2006). This work is based on the analysis of the large subunit of the nuclear ribosome LSU or 28S sampled from 31 diverse medusozoans. The tree is rooted with anthozoans. Numbers in parentheses indicate the number of species sampled and the bar indicates 0.05 substitutions per site used in the model of nucleotide evolution

3.4.3 M. kingi final annotation

Table 3.2 and Table 3.3 show homologous sequences to cubozoans using BLASTX with a threshold of 1e-5. A very small percentage of ESTs comprise full length coding regions and a significant number of ESTs displayed extremely weak similarity to public sequences.

Table 3.2 Summary of the annotation inferred by BLASTX

Last assessed July 31st 2008. The overall accuracy for the functional annotation using https://fungalgenome.concordia.ca/tools/TargetIdentifier.html is > 90% (Min *et al.* 2005).

Total number of sequences		***1114***
number of sequences with no hit		229
number of sequences with hit		341
Sequences with e-Values less than the threshold		544
full-length sequences		70
short full-length sequences		12
possible full-length sequences		0
ambiguous sequences		0
partial sequences (5'-sequenced partial)		49
3'-sequenced partial sequences		33
full-length & ORF completely sequenced sequences		37
number of hits with e-values <= 1e-100		14
number of hits with e-values <= 1e-50 and > 1e-100	Homologous to *M. kingi*	37
number of hits with e-values <= 1e-30 and > 1e-50	Highly similar to	36
number of hits with e-values <= 1e-10 and > 1e-30	Similar to	60
number of hits with e-values <= 1e-5 and > 1e-10	Weakly similar to	24
number of hits with e-values <= 0.1 and > 1e-5	Very weakly similar to	25
number of hits with e-values > 0.1	Extremely weakly similar to *M. kingi*	145

Table 3.3 Annotated ESTs using a stringent threshold of 1e-5

The best match out of ten on BLASTX similarity comparisons. The left column gives gene annotation, the database identifiers, e value, species similarity, followed by *M. kingi* EST cluster number, mRNA sequence length and cDNA strand and orientation of encoded product. Importantly in this table, a considerable number of clusters is seen showing higher similarities to vertebrate than to invertebrate sequences, which has been observed in the coral *A. millepora* (Kortschak *et al.* 2003), the anemone *N. vectensis* and in *Hydra.* Eight clusters are similar to bacterial genes. These genes are unlikely to be contamination, due to the presence of other non metazoan genes that have been reported in genomes of anemones and corals (Technau 2005).

	Proteases some proteases		Metal binding proteins		Important in development
	Protease inhibitor		Transcription Factor		Innate immune system
	Toxins and putative toxins				

ANNOTATION	Gene DB ID	E Value	Species similarity	Mk CLUSTER	mRNA	cDNA strand
Cyclophilin A	gb|AAR11779.1	4e-52	*Chlamys farrer*	72	FL	antisense
Actin, non-muscle 6.2	(P17126)	0	Hydrozoa	215	FL	sense complete
Soma ferritin (EC 1.16.3.1)	(P42577)	1e-58	FRIS_LYMST	69	FL	sense complete
Elongation factor 1-alpha (EF-1-alpha)	(P05303)	0	FLy	1038	FL	sense complete
ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.3.14) (ATP synthase proteolipid P2) (ATPase protein 9) (ATPase	(Q5RAP9)	3e-18		206	FL	antisense
Gamma-butyrobetaine dioxygenase (EC 1.14.11.1) (Gamma-butyrobetaine,2-oxoglutarate dioxygenase) (Gamma-butyrobetaine	(O75936)	1e-58	Human	25	FL	sense complete
Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-)	(P40313)	1e-48	Human	212	FL	sense complete
Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-)	(P40313)	3e-42	Human	101	FL	sense complete
Myosin light chain 1, slow-twitch muscle A isoform (MLC1sa) (Alkali)	(P14649)	7e-28	Human	68	FL	sense complete
Inhibitor of carbonic anhydrase precursor	(Q29545)	3e-48	Pig	216	FL	sense partial
110 kDa antigen (PK110) (Fragment)	(P13813)	6e-07	110KD_PLAKN	76	FL	sense complete
Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase)	(P00450)	1e-33	Human	198	FL	antisense
Succinate dehydrogenase	(Q9CQA3)	1e-113	Mouse	46	FL	antisense
Myosin-10 (Myosin heavy chain, nonmuscle IIb) (Nonmuscle myosin heavy chain IIb) (NMMHC II-b) (NMMHC-IIB) (Cellular myosin	(Q27991)	1e-47		106	FL	sense partial
Neuropeptide FF receptor 2 (Neuropeptide G-protein coupled receptor) (G-protein coupled receptor 74) (G-protein coupled receptor	(Q9Y5X5)	4e-08	Human	62	FL	sense complete
Probable COP9 signalosome complex subunit 7 (Signalosome subunit 7)	(Q94261)	7e-40	Worm	127	FL	sense complete
Cystathionine gamma-lyase (EC 4.4.1.1) (Gamma-cystathionase)	(P32929)	2e-67	Human	154	FL	sense partial
Retrovirus-related Pol polyprotein from type I retrotransposable element R2	(Q05118)	3e-07		173	FL	antisense
2-amino-3-carboxymuconate-6-semialdehyde decarboxylase (EC 4.1.1.45)	(Q8R5M5)	1e 136		110	FL	sense complete
5-formyltetrahydrofolate cyclo-ligase (EC 6.3.3.2) (5,10-methenyl-tetrahydrofolate synthetase) (Methenyl-THF synthetase) (MTHFS)	(P80405)	1e-36	Rabitt	126	FL	antisense
Soma ferritin (EC 1.16.3.1)	(P42577)	1e-58		152	FL	sense complete
Tagatose-1,6-bisphosphate aldolase agaY (EC 4.1.2.-) (TBPA)	(P0AB74)	1e 161	Bacteria	174	FL	sense complete
Membrane-bound transcription factor site 1 protease precursor (EC 3.4.21.-) (S1P endopeptidase) (Site-1 protease) (Subtilisin/kexin	(Q9WTZ3)	3e-10	Rat	185	FL	antisense
Betaine--homocysteine S-methyltransferase (EC 2.1.1.5)	(Q93088)	1e-158	Human	202	FL	sense complete
ES1 protein homolog, mitochondrial precursor	(Q9D172)	3e-23	Mouse	205	FL	antisense
Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)	(Q90380)	1e-56		6	FL	sense complete
Glycine cleavage system H protein, mitochondrial precursor	(Q9N121)	2e-29	Rabitt	51	FL	sense complete
Kinesin-like protein KIF13A	(Q9EQW7)	1e-10	Mouse	63	FL	sense partial
Craniofacial development protein 2 (p97 bucentaur protein)	(O02751)	2e-09	Bovine	452	FL	sense partial
Retrovirus-related Pol polyprotein from type I retrotransposable element R2	(Q05118)	5e-07		455	FL	sense partial
Dynein light chain 2, cytoplasmic (Dynein light chain LC8-type 2)	(Q78P75)	7e-42	Rat	479	FL	sense complete
Heme-binding protein 2 (Protein SOUL) (Placental protein 23) (PP23)	(Q9Y5Z4)	3e-17	Human	498	FL	sense complete
Alpha-2-macroglobulin precursor (Alpha-2-M)	(Q5R4N8)	6e-28		561	FL	sense partial
Cytochrome c oxidase copper chaperone	(Q6J3Q7)	2e-17		568	FL	sense complete
FRAS1-related extracellular matrix protein 2 precursor (ECM3 homolog)	(Q5SZK8)	2e-41	Human	581	FL	antisense
Heparin cofactor II precursor (HC-II) (Protease inhibitor leuserpin 2)	(P49182)	2e-10	Mouse	598	FL	sense partial
Craniofacial development protein 2 (p97 bucentaur protein)	(Q588U8)	1e-06		622	FL	antisense
U3 small nucleolar ribonucleoprotein protein IMP3 (U3 snoRNP protein IMP3)	(Q921Y2)	6e-08	Mouse	645	FL	antisense
Retrovirus-related Pol polyprotein from type II retrotransposable element R2DM	(P16423)	2e-06	FLY	697	FL	antisense
Craniofacial development protein 2 (p97 bucentaur protein)	(O02751)	3e-17	Bovine	706	FL	antisense
Hemagglutinin/amebocyte aggregation factor precursor (18K-LAF)	(Q01528)	4e-07		718	FL	sense complete
Prolyl 4-hydroxylase alpha-1 subunit precursor (EC 1.14.11.2) (4-PH alpha-1) (Procollagen-proline,2-oxoglutarate-4-dioxygenase alpha-	(Q10576)	1e-09	Worm	722	FL	sense partial
Eukaryotic translation initiation factor 2 subunit 3, Y-linked (Eukaryotic translation initiation factor 2 gamma subunit, Y-linked) (eIF-2-	(Q9Z0N2)	1e-12	Mouse	733	FL	antisense
Chitotriosidase-1 precursor (EC 3.2.1.14) (Chitinase-1)	(Q13231)	6e-06	Human	757	FL	sense complete
Craniofacial development protein 2 (p97 bucentaur protein)	(O02751)	6e-18	Bovine	767	FL	sense partial
Glycogen synthase kinase-3 alpha (EC 2.7.1.37) (GSK-3 alpha) (Factor A) (FA)	(P18265)	7e-87	Rat	775	FL	sense partial

Table 3.3 (*continued)* Annotated ESTs using a stringent threshold of 1e-5

ANNOTATION	Gene Data Base	E Value	Species similarity	CLUSTER ID	mRNA	cDNA strand
Peptidyl-prolyl cis-trans isomerase, mitochondrial precursor (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin F)	(P29117)	8e-52	Rat	72	SHORT (42) FL	antisense
Blastula protease 10 precursor (EC 3.4.24.-)	(P42674)	1e-35		131	SHORT (32) FL	sense partial
Mitochondrial glutamate carrier 1 (Glutamate/H(+) symporter 1) (Solute carrier family 25 member 22)	(Q9H936)	9e-73	Human	140	SHORT (60) FL	sense complete
Toxin A precursor (CqTX-A)	(P58762)	7e-10	Box jellyfish	547	SHORT (93) FL	sense partial
F-actin capping protein alpha-2 subunit (CapZ 36/32) (Beta-actinin subunit I)	(P28497)	7e-71	Chick	574	SHORT (94) FL	sense complete
Probable reductase (EC 1.1.-.-)	(P22045)	7e-20		601	SHORT (98) FL	sense complete
ADP-ribosylation factor-like protein 6	(O88848)	1e-72	Mouse	625	SHORT (60) FL	sense complete
Cytochrome c1, heme protein, mitochondrial (Cytochrome c-1)	(P00125)	2e-65	Bovine	644	SHORT (43) FL	antisense
EGF-like domain-containing protein 3 precursor (Multiple EGF-like domain protein 3) (Multiple epidermal growth factor-like domains	(O88281)	5e-09	Rat	652	SHORT (24) FL	antisense
Fibrinogen-like protein A precursor (FREP-A)	(P19477)	1e-36		799	SHORT (14) FL	sense partial
Carboxypeptidase A6 precursor (EC 3.4.17.1)	(Q5U901)	3e-29	Mouse	803	SHORT (15) FL	sense partial
Protein arginine N-methyltransferase 4 (EC 2.1.1.-) (Heterogeneous nuclear ribonucleoprotein methyltransferase-like protein 4)	(Q6PAK3)	2e-92	Mouse	947	SHORT (21) FL	sense partial
Retrovirus-related Pol polyprotein from type I retrotransposable element R2	(Q03274)	9e-08		970	SHORT (54) FL	antisense
52 kDa repressor of the inhibitor of the protein kinase (p58IPK-interacting protein) (58 kDa interferon-induced protein kinase-interacting	(O43422)	1e-08	Human	1015	SHORT (94) FL	sense partial
Protein ea10	(P68658)	3e-18		239	SHORT (95) FL	antisense
Cytochrome bd-II oxidase subunit 1 (EC 1.10.3.-) (Cytochrome bd-II oxidase subunit I)	(P26459)	1e-117	Bacteria	242	SHORT (51) FL	sense partial
Homocysteine S-methyltransferase 1 (EC 2.1.1.10) (S-methylmethionine:homocysteine methyltransferase 1) (SMM:Hcy S-	(Q9SDL7)	3e-48		316	SHORT (86) FL	sense partial
Elongation factor 1-alpha 2 (EF-1-alpha-2) (Elongation factor 1 A-2) (eEF1A-2) (Statin S1)	(Q71V39)	3e-75	Rabit	1039	Partial	sense complete
Transforming growth factor-beta-induced protein ig-h3 precursor (Beta ig-h3) (Kerato-epithelin) (RGD-containing collagen-associated	(O11780)	2e-36	Pig	213	Partial	sense complete
Apolipophorins precursor	(Q9U943)	2e-24		34	Partial	sense complete
Oviduct-specific glycoprotein precursor (Oviductal glycoprotein) (Oviductin) (Estrogen-dependent oviduct protein)	(Q62010)	4e-07	Mouse	207	Partial	sense complete
Parathyroid hormone-related protein precursor (PTH-rP) (PTHrP)	(Q9GLC7)	3e-15	Rabit	3	Partial	sense partial
Collagen alpha-1(II) chain precursor (Fragment)	(P02460)	2e-49	Chick	94	Partial	sense complete
Retrovirus-related Pol polyprotein from type I retrotransposable element R2	(Q03274)	5e-06		40	Partial	sense partial
Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8)	(Q90473)	E-121		218	Partial	sense complete
Zinc metalloproteinase nas-15 precursor (EC 3.4.24.21) (Nematode astacin 15)	(P55115)	1e-35	Worm	105	Partial	sense partial
Hemicentin-1 precursor (Fibulin-6) (FIBL-6)	(Q96RW7)	5e-21	Human	91	Partial	sense partial
Collagen alpha-1(II) chain precursor (Fragment)	(P02460)	7e-44	Chick	176	Partial	sense complete
26S protease regulatory subunit 8 (18-56 protein)	(P54814)	E-154		100	Partial	sense complete
Protein disulfide-isomerase A4 precursor (EC 5.3.4.1) (Protein ERp-72) (ERp72)	(P08003)	1e-82	Mouse	130	Partial	sense complete
Programmed cell death 6-interacting protein (Signal transduction protein Xp95)	(Q9W6C5)	9e-27	Frog	141	Partial	sense complete
Interleukin enhancer-binding factor 3 (Nuclear factor of activated T-cells 90 kDa) (NF-AT-90) (Double-stranded RNA-binding protein	(Q12906)	5e-51	Human	168	Partial	sense partial
Succinate dehydrogenase	(Q8K2B3)	2e-55	Mouse	178	Partial	sense complete
Protein PRY1 precursor (Pathogen related in Sc 1)	(P47032)	2e-21	Yeast	197	Partial	sense complete
Craniofacial development protein 2 (p97 bucentaur protein)	(Q588U8)	2e-18		200	Partial	sense complete
Hemocytin precursor (Humoral lectin)	(P98092)	1e-07		23	Partial	sense partial
Zinc finger protein Kr18 (HKr18)	(Q9HCG1)	1e-34	Human	50	Partial	sense partial
Serine carboxypeptidase precursor (EC 3.4.16.-)	(P32826)	3e-17		53	Partial	sense complete
Caspase-8 precursor (EC 3.4.22.-) (CASP-8)	(O89110)	2e-20	Mouse	457	Partial	sense partial
ATP synthase gamma chain, mitochondrial precursor (EC 3.6.3.14)	(O01666)	1e-21	FLy	481	Partial	sense complete
ATP synthase delta chain, mitochondrial precursor (EC 3.6.3.14)	(Q9D3D9)	3e-35	Mouse	502	Partial	sense complete
NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-75Kd) (CI-75Kd)	(P28331)	4e-79	Human	508	Partial	sense partial
Chymotrypsin B (EC 3.4.21.1)	(P80646)	5e-12		521	Partial	sense complete
Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3) (ALDH class 2) (ALDH1) (ALDH-E2)	(P81178)	1e-67		559	Partial	sense partial

Table 3.3(*continued)* Annotated ESTs using a stringent threshold of 1e-5

Coiled-coil-helix-coiled-coil-helix domain-containing protein 2 (HCV NS2 trans-regulated protein) (NS2TP)	(Q9Y6H1)	7e-19	Human	563	Partial	sense complete
Puromycin-sensitive aminopeptidase (EC 3.4.11.-) (PSA)	(Q11011)	1e-22	Mouse	569	Partial	sense complete
Serine protease hepsin (EC 3.4.21.-)	(O35453)	8e-38	Mouse	578	Partial	sense complete
Eukaryotic translation initiation factor 3 subunit 5 (eIF-3 epsilon) (eIF3 p47 subunit) (eIF3f)	(O00303)	2e-46	Human	595	Partial	sense complete
Glutamate decarboxylase 1 (EC 4.1.1.15) (Glutamate decarboxylase, 67 kDa isoform) (GAD-67) (67 kDa glutamic acid decarboxylase)	(Q99259)	5e-91	Human	597	Partial	sense partial
DnaJ homolog subfamily C member 1	(Q96KC8)	4e-10	Human	611	Partial	sense complete
EGF-like repeat and discoidin I-like domain-containing protein 3 precursor (EGF-like repeats and discoidin I-like domains protein 3)	(O43854)	1e-13	Human	615	Partial	sense partial
Extracellular matrix protein FRAS1 precursor	(Q80T14)	9e-13	Mouse	620	Partial	sense partial
Copia protein (Gag-int-pol protein)	(P04146)	2e-07	FLy	660	Partial	sense complete
NEDD8 precursor (Ubiquitin-like protein Nedd8) (Neddylin) (Protein NED-8)	(Q93725)	5e-14	Worm	665	Partial	sense complete
Probable urocanate hydratase (EC 4.2.1.49) (Urocanase) (Imidazolonepropionate hydrolase)	(Q8VC12)	1e-67	Mouse	676	Partial	sense complete
Collagen alpha-1(I) chain precursor	Q9XSJ7)	2e-26		694	Partial	sense partial
Endoplasmin precursor (Heat shock 108 kDa protein) (HSP108) (HSP 108) (Transferrin-binding protein)	(P08110)	3e-51	Chick	696	Partial	sense partial
Adenylyl cyclase-associated protein 1 (CAP 1)	(P40124)	3e-48	Mouse	699	Partial	sense partial
Eukaryotic translation initiation factor 3 subunit 6 (eIF-3 p48) (eIF3e) (Mammary tumor-associated protein INT-6) (Viral integration site	(P60229)	E-112	Mouse	708	Partial	sense partial
Ubiquitin-conjugating enzyme spm2 (Ubiquitin-conjugating enzyme variant MMS2 homolog) (UEV MMS2)	(O74983)	2e-07	Yeast	715	Partial	sense complete
Polyadenylate-binding protein 1 (Poly(A)-binding protein 1) (PABP 1)	(P29341)	2e-50	Mouse	727	Partial	sense partial
Inositol-3-phosphate synthase (EC 5.5.1.4) (Myo-inositol-1-phosphate synthase) (MI-1-P synthase) (IPS)	(Q9S7U0)	E-104		739	Partial	sense partial
Heat shock 70 kDa protein cognate 5	(P29845)	8e-21	FLy	745	Partial	sense complete
Eukaryotic translation initiation factor 3 subunit 3 (eIF-3 gamma) (eIF3 p40 subunit) (eIF3h)	(Q91WK2)	2e-38	Mouse	750	Partial	sense complete
Aureobasidin A resistance protein homolog	(Q10142)	2e-24	Yeast	755	Partial	sense partial
Protein KIAA1199 homolog precursor (Protein 12H19.01.T7)	(Q8BI06)	2e-09	Mouse	758	Partial	sense partial
Elastase-1 precursor (EC 3.4.21.36)	(Q9UNI1)	1e-30	Human	770	Partial	sense complete
Peptide methionine sulfoxide reductase msrB (EC 1.8.4.6)	(Q8DJK9)	2e-41		827	FL	sense complete
Epididymal secretory protein E1 precursor (Niemann Pick type C2 protein homolog) (16.5 kDa secretory protein)	(Q9DGJ3)	1e-18		845	FL	sense complete
Osmoregulated proline transporter (Sodium/proline symporter)	(O06493)	1e-07		874	FL	sense partial
Apoptosis inhibitor FKSG2	(Q9HAU6)	2e-10	Human	877	FL	sense complete
GTP-binding nuclear protein GSP2/CNR2	(P32836)	7e-45	Yeast	893	FL	sense partial
Ferritin-1 heavy chain (EC 1.16.3.1)	(P25319)	1e-09	Yeast	895	FL	sense partial
Cytochrome c	(Q6QLW4)	3e-44		903	FL	sense partial
Jerky protein	(Q60976)	9e-06	Mouse	953	FL	sense partial
UPF0203 protein 15E1.1	(Q9D8Z2)	4e-12	Mouse	961	FL	sense complete
NADH-ubiquinone oxidoreductase 19 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-19KD) (CI-19KD) (Complex I-PGIV) (CI-	(P51970)	4e-32	Human	973	FL	sense complete
Elastase-2A precursor (EC 3.4.21.71) (Elastase-2)	(P08419)	4e-11	Pig	982	FL	antisense
Galactokinase (EC 2.7.1.6) (Galactose kinase)	(Q83M01)	1e-08		995	FL	antisense
Probable ATPase inhibitor B0546.1, mitochondrial precursor	(O44441)	3e-06	Worm	999	FL	sense complete
Myosin light chain 1, skeletal muscle isoform (MLC1F) (A1 catalytic) (Alkali myosin light chain 1)	(P05976)	4e-29	Human	1001	FL	sense complete
Glyoxylate reductase (EC 1.1.1.26) (Glycolate reductase)	(Q8U3Y2)	2e-30		1020	FL	sense partial
Probable general secretion pathway protein E (Type II traffic warden ATPase)	(P45759)	2e-87	Bacteria	227	FL	antisense
Curli production assembly/transport component csgF precursor	(P0AEA1)	2e-62		228	FL	antisense
Probable csgAB operon transcriptional regulatory protein	(P52106)	9e-77	Bacteria	229	FL	sense partial
Molybdopterin biosynthesis protein moeB	(P12282)	7e-10	Bacteria	243	FL	sense partial
6-phosphogluconolactonase (EC 3.1.1.31) (6-P-gluconolactonase)	(Q83LP7)	8e-15		245	FL	antisense
Exonuclease (EC 3.1.11.3)	(P03697)	1e-95		249	FL	sense partial
Protein ea	(P68658)	2e-17		253	FL	sense partial
Green FLuorescent protein	(P42212)	1e-38	Aequorea victoria	283	FL	antisense

Table 3.3(*continued)* Annotated ESTs using a stringent threshold of 1e-5

ANNOTATION	Gene DB ID	E Value	Species similarity	CLUSTER	mRNA	cDNA strand
Probable ATP-dependent RNA helicase DDX17 (EC 3.6.1.-) (DEAD box protein 17)	(Q501J6)	7e-85	Mouse	791	Partial	sense partial
EGF-like domain-containing protein 3 precursor (Multiple EGF-like domain protein 3) (Multiple epidermal growth factor-like domains	(O88281)	1e-22	Rat	804	Partial	sense partial
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform (PP2A, subunit A, PR65-alpha isoform) (PP2A,	(P54612)	5e-79	Pig	830	Partial	sense complete
Copia protein (Gag-int-pol protein)	(P04146)	2e-40	FLy	840	Partial	sense partial
Heat shock protein HSP 90-alpha	(P11501)	8e-54	Chick	862	Partial	sense complete
Zinc finger protein sens (Protein senseless)	(Q9N658)	4e-19	FLy	878	Partial	sense partial
Succinate dehydrogenase	(Q920L2)	E-129	Rat	891	Partial	sense partial
Ubiquitin--protein ligase hyd (EC 6.3.2.-) (Protein hyperplastic discs)	(P51592)	5e-13	FLy	921	Partial	sense complete
Proteasome subunit beta type 1-A (EC 3.4.25.1) (20S proteasome beta 6 subunit A) (B6-A)	(Q9IB84)	6e-36		927	Partial	sense complete
Apolipophorins precursor	(Q9U943)	2e-21		990	Partial	sense partial
60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1)	(P37578)	8e-06		1022	Partial	sense partial
Nonspecific lipid-transfer protein, mitochondrial precursor (EC 2.3.1.176) (Propanoyl-CoA C-acyltransferase) (NSL-TP) (Sterol carrier	(O62742)	5e-70	Rabit	1026	Partial	sense partial
Guanine nucleotide-binding protein beta subunit 2-like 1 (Receptor of activated protein kinase C) (Schulte *et al.*)	(O42249)	E-103		1027	Partial	sense complete
Putrescine transport system permease protein potI	(P0AFL1)	1e-07	Bacteria	226	Partial	sense complete
Quinolinate synthetase A	(P11458)	1e-56	Bacteria	246	Partial	sense partial
Arginine kinase (EC 2.7.3.3) (AK)	(Q9NH48)	4e-18		250	Partial	sense partial
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12) (E2) (Dihydrolipoamide	(P08461)	1e-32	Rat	298	Partial	sense partial
52 kDa repressor of the inhibitor of the protein kinase (p58IPK-interacting protein) (58 kDa interferon-induced protein kinase-interacting	(Q9CUX1)	2e-13	Mouse	60	3'-PARTIAL seq	antisense
ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14)	(Q5ZLC5)	E-158	Chick	35	3'-PARTIAL seq	antisense
Elongation factor 2 (EF-2)	(Q90705)	1e-76	Chick	107	3'-PARTIAL seq	antisense
Cystathionine gamma-lyase (EC 4.4.1.1) (Gamma-cystathionase)	(P32929)	5e-59	Human	124	3'-PARTIAL seq	antisense
Myosin-10 (Myosin heavy chain, nonmuscle IIb) (Nonmuscle myosin heavy chain IIb) (NMMHC II-b) (NMMHC-IIB) (Cellular myosin	(Q9JLT0)	1e-32	Rat	75	3'-PARTIAL seq	antisense
BMP-binding endothelial regulator protein precursor (Crossveinless-2 protein) (mCV2)	(Q8CJ69)	4e-10	Mouse	29	3'-PARTIAL seq	antisense
Mitochondrial inner membrane protein (Mitofilin)	(Q8CAQ8)	9e-28	Mouse	114	3'-PARTIAL seq	antisense
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit precursor (EC 2.4.1.119) (Oligosaccharyl transferase 48	(O54734)	7e-80	Mouse	150	3'-PARTIAL seq	antisense
Hemicentin-1 precursor (Fibulin-6) (FIBL-6)	(Q96RW7)	1e-36	Human	160	3'-PARTIAL seq	antisense
Chaperone protein htpG (Heat shock protein htpG) (High temperature protein G)	(P58480)	2e-45		194	3'-PARTIAL seq	antisense
Coronin-6 (Clipin E)	(Q920M5)	1e-23	Mouse	47	3'-PARTIAL seq	antisense
NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-75Kd) (CI-75Kd)	(Q91VD9)	1e-58	Mouse	57	3'-PARTIAL seq	antisense
Caspase-3 precursor (EC 3.4.22.-) (CASP-3) (Apopain) (Cysteine protease CPP32) (xCPP32)	(P55866)	8e-17	Frog	456	3'-PARTIAL seq	antisense
Heat shock cognate 71 kDa protein	(P47773)	1e-09		477	3'-PARTIAL seq	antisense
Murinoglobulin-2 precursor (MuG2)	(P28666)	3e-14	Mouse	486	3'-PARTIAL seq	antisense
Retinal dehydrogenase 1 (EC 1.2.1.36) (RalDH1) (RALDH 1) (Aldehyde dehydrogenase family 1 member A1) (Aldehyde	(P51977)	5e-62	Sheep	560	3'-PARTIAL seq	antisense
Trifunctional enzyme alpha subunit, mitochondrial precursor (TP-alpha)	(Q64428)	3e-98	Rat	589	3'-PARTIAL seq	antisense
Ubiquinone biosynthesis protein COQ9-A, mitochondrial precursor	(Q3B8B2)	2e-27	Frog	662	3'-PARTIAL seq	antisense
Eukaryotic translation initiation factor 2 subunit 1 (Eukaryotic translation initiation factor 2 alpha subunit) (eIF-2-alpha) (EIF-2alpha)	(P68101)	6e-15	Rat	705	3'-PARTIAL seq	antisense
LINE-1 reverse transcriptase homolog	(P08548)	5e-10		763	3'-PARTIAL seq	antisense
Major head protein (gpE) (Major coat protein)	(P03713)	7e-28		789	3'-PARTIAL seq	antisense
Coactosin-like protein	(Q14019)	5e-15	Human	211	SHORTH FL (69)	sense complete
CCAAT/enhancer-binding protein beta (C/EBP beta) (Interleukin-6-dependent-binding protein) (IL-6DBP) (Liver-enriched	(P28033)	1e-14	Mouse	56	SHORT FL (28)	antisense
ATP synthase mitochondrial F1 complex assembly factor 2, mitochondrial precursor	(Q91YY4)	3e-26	Mouse	901	3'-PARTIAL seq	antisense
Leukocyte cell-derived chemotaxin 2 precursor (Chondromodulin II) (ChM-II)	(O88803)	5e-20	Mouse	1012	3'-PARTIAL seq	antisense
FKBP12-rapamycin complex-associated protein (FK506-binding protein 12-rapamycin complex-associated protein 1) (Rapamycin target	(P42346)	1e-18	Rat	1034	3'-PARTIAL seq	antisense
Dimethyl sulfoxide reductase precursor (EC 1.8.99.-) (DMSO reductase) (DMSOR)	(Q57366)	4e-40		230	3'-PARTIAL seq	antisense
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial precursor (EC 2.3.1.12)	(P10515)	3e-30	Human	299	3'-PARTIAL seq	antisense
Sulfide:quinone oxidoreductase, mitochondrial precursor (EC 1.-.-.-) (Heavy metal tolerance protein 2) (Cadmium resistance protein 1)	(O94284)	6e-06	Yeast	315	3'-PARTIAL seq	antisense
110 kDa antigen (PK110) (Fragment)	(P13813)	1e-09		77	Ambiguous	sense complete

3.4.4 Open reading frames, prediction of secreted, non-classical and leaderless secreted proteins

Since cnidarians use nematocysts to store toxic components, and these toxic components are expressed via a secretory pathway, a search for secreted proteins and signal peptides was conducted, typically between 15 and 40 amino acids long and with a basic charge. This feature was extensively explored to identify proteins that may be confined to nematocysts. Figure 3.4 shows the results of these analyses.

The 1,039 DNA clusters (818 singlets) were submitted to the ORF predictor at https://fungalgenome.concordia.ca, taking into consideration the six possible ORFs. This resulted in 982 conceptual amino acid sequences (not necessarily beginning with an initiating methione) that were used for further searches.

To identify the presence or absence and location of signal peptide cleavage sites, the amino acid sequences were submitted to the Centre for Biological Sequence Analysis BioCentrum-DTU Technical University of Denmark (SignalP 3.0 Server) using neural networks (NN) and hidden Markov models (HMM) trained on eukaryotes (Bendtsen *et al.* 2004a; Nielsen *et al.* 1997). The entire data set was accepted for the program, retrieving 176 clusters predicted to contain signal peptides in both methods. In addition, open reading frames were submitted to Secretome with a neural network score threshold of 0.5. This web based program produced *ab initio* predictions of non-classical *i.e.* protein secretion not triggered by a signal peptide. Only 802 predicted peptides were accepted for the program [180 amino acid sequences (18.32%) were not accepted because they were less than 45 amino acids long]. The method was informative because it queries a large number of other feature prediction servers to obtain information on various post-translational and localization aspects of the protein sequence, which are integrated into a final secretion prediction. 793 protein sequences produced results; 142 of those were considered non-secreted proteins, 143 predicted to be secreted by both SignalP and Secretome server, 508 predicted proteins were considered likely to be products of non-classical or leaderless secretion pathways according to neural network scores (0.5-0.98).

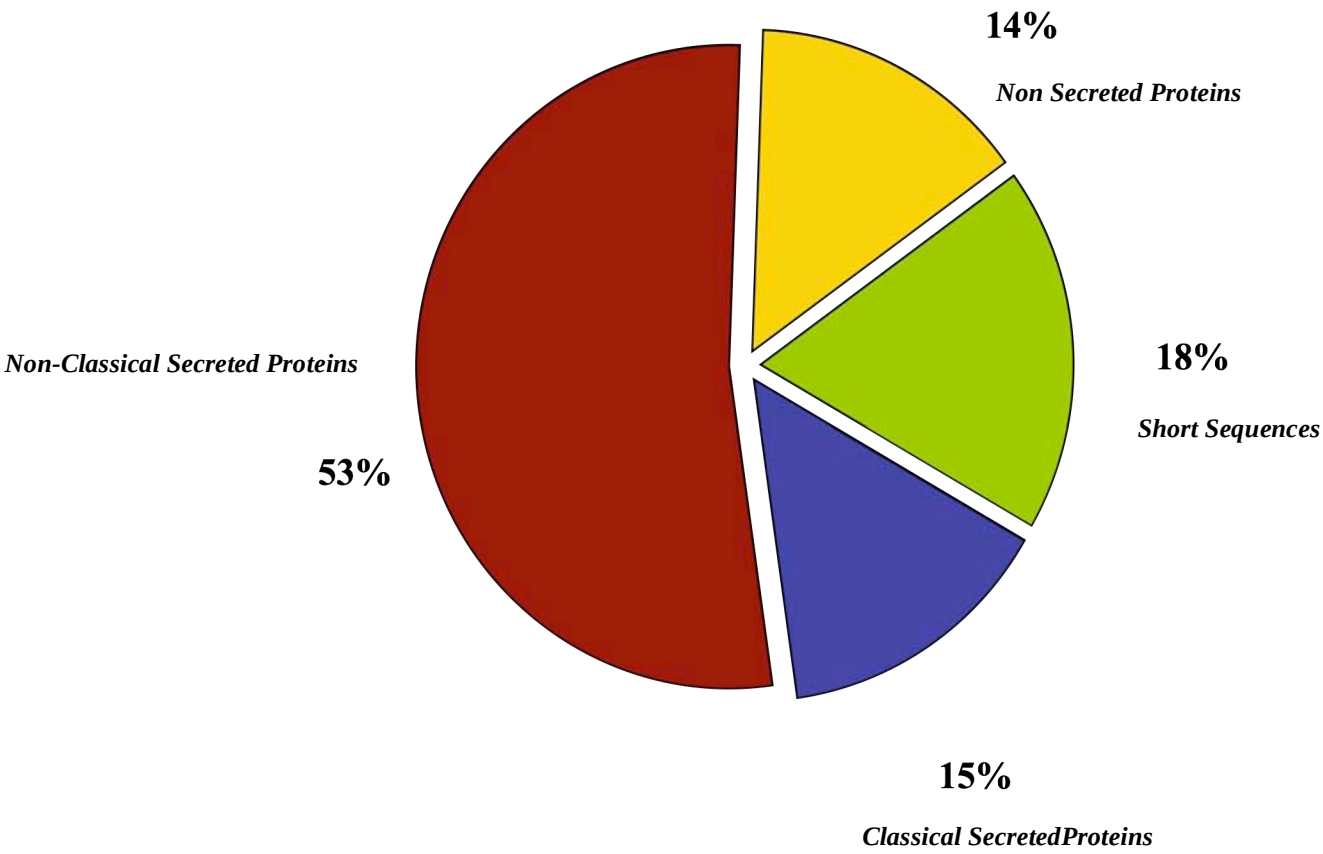


Figure 3.4 *M. kingi* predicted open reading frame distribution based on the presence or absence of a signal peptide.

Amino acid sequences of 802 predicted open reading frames were used to predict secreted proteins. Secreted proteins are represented in blue, non-secreted proteins represented in yellow, proteins predicted to undertake a non-classical secreted pathway are indicated in red and protein sequences that were too short to accurately produce a prediction are indicated in green.

Predictions of signal cleavage sites were considered to be more accurate when a secreted protein was detected by both SignalP (NN and HMM) and Secretome (NN). Sequences that did not match the criteria were analysed manually and using MacVector. It was found that accurate predictions depended upon completeness of the cDNA sequences, however because these data are ESTs a vast majority were not full length. Although a larger data set is needed to identify secreted products with more certainty, we were able to find secreted proteins from *M. kingi*, believed to be located in the nematocyst. A representative of this identified group of genes is the nematocyst-specific minicollagen-I (accession number EU753870). In cnidarians this gene is believed to be a major component of the nematocyst wall (David *et al.* 2008; Engel *et al.* 2001).

3.4.5 Classification by functional class and gene ontology classification of M. kingi *ESTs.*

To categorize transcripts by putative function and ontology classification, the deduced amino acid sequences were submitted to ProtFun 2.2 Server (http://www.cbs.dtu.dk/services/ProtFun/). The resulting functional and gene ontology predictions are summarized in Figure 3.5 and Table 3.4. Results indicate that numerous transcripts encode proteins involved in energy metabolism and transport as well as cell envelope proteins. According to gene ontology (GO) many of the transcripts belong to growth factor and immune response families (see Figure 3.11). However when groups of interest were analysed in detail low prediction accuracy was detected. Furthermore considerable amounts of data were rejected from the ProtFun 2.2 Server, likely to be the result of short products. In spite of these issues most of the data was manually inspected and important transcripts were identified.

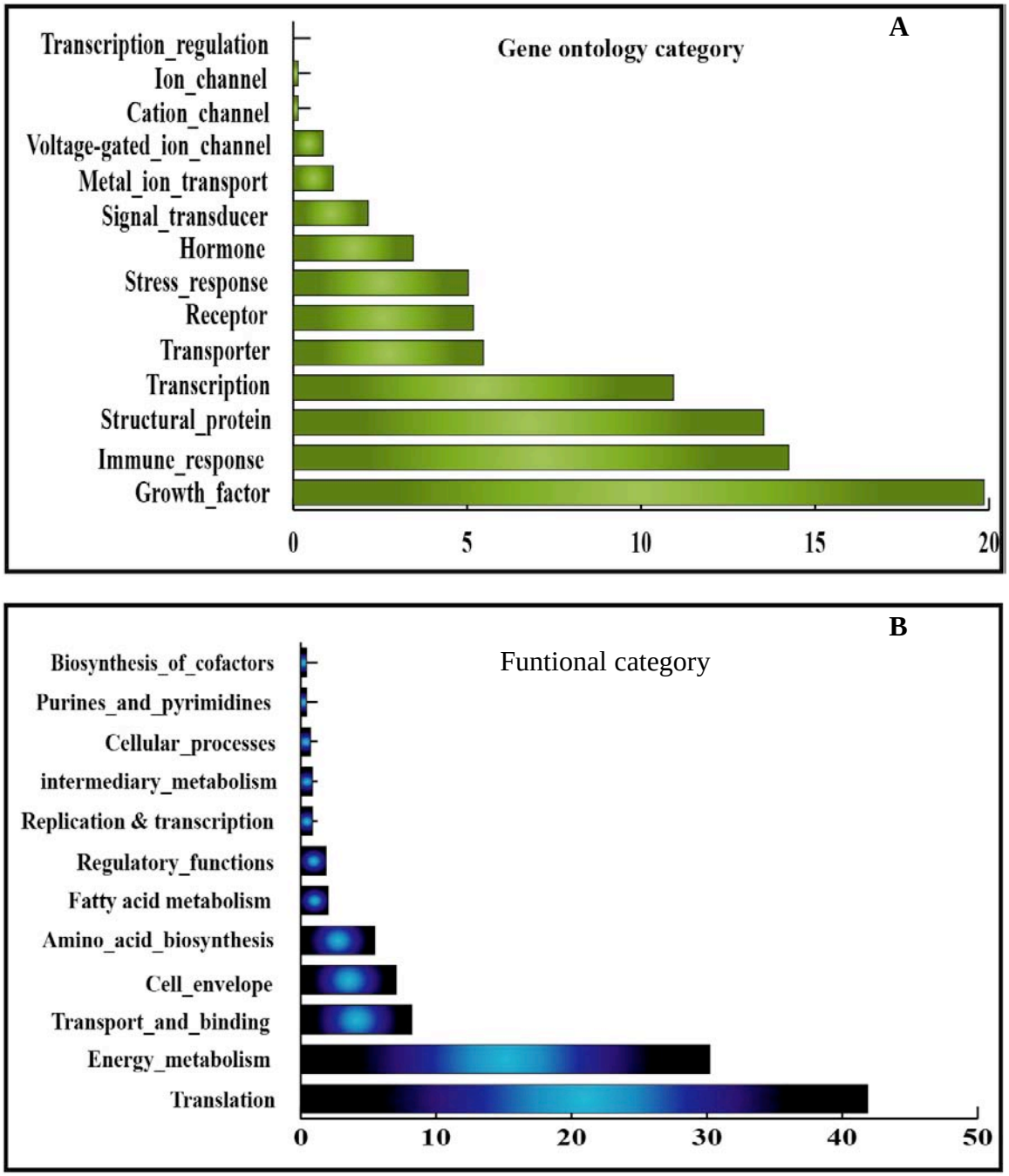


Figure 3.5 Functional classification and gene ontology category

A and B show predictions for gene ontology and functional classification, respectively. The data are represented as percentages using a total of 695 translated transcripts. (cbs.dtu.dk/services/ProtFunt).

Table 3.4 Characterization of ESTs: Prediction of cellular role, enzyme class and gene ontology category

Secreted, non-classical secreted and non-secreted proteins were identified. 802 proteins were submitted for analysis, 695 were accepted by ProtFun 2.2 (Jensen *et al.* 2002; Jensen *et al.* 2003).

	Whole Set (695 retrieved queries)	*Secreted*	*Non Classical & leaderless Secreted*	*Non Secreted*
Functional category				
Amino_acid_biosynthesis	31	4	27	7
Biosynthesis_of_cofactors	2	0	2	1
Cell_envelope	47	**35**	12	2
Cellular_processes	5	0	5	0
Central_intermediary_metabolism	4	0	4	2
Energy_metabolism	164	46	118	46
Fatty_acid_metabolism	13	1	12	1
Purines_and_pyrimidines	2	1	1	1
Regulatory_functions	11	0	11	2
Replication_and_transcription	3	0	3	3
Translation	246	4	242	45
Transport_and_binding	53	34	19	4
*	**478 queries falled into a category** (68.77% from total)	125	456	114
Enzyme/nonenzyme				
Enzyme	307	73	234	79
Nonenzyme	270	48	222	43
*	**100% of the sequences were classified into a class**	**121**	**456**	**122**
Enzyme class				
Oxidoreductase	0	3	0	0
Transferase	0	0	0	0
Hydrolase	0	1	0	0
Lyase	36	10	14	12
Isomerase	23	4	11	8
Ligase	59	2	1	3
*	**118 predicted to fall into an enzime class** (Only 16.97 %)	20	26	23
Gene Ontology category				
Signal_transducer	15	10	3	2
Receptor	30	15	15	6
Hormone	21	**14**	7	3
Structural_protein	82	13	69	12
Transporter	24	14	10	14
Ion_channel	0	0	0	1
Voltage-gated_ion_channel	6	1	5	0
Cation_channel	1	0	1	0
Transcription	75	1	74	1
Transcription_regulation	1	0	0	1
Stress_response	33	**18**	15	2
Immune_response	85	**11**	74	14
Growth_factor	94	5	89	44
Metal_ion_transport	8	8	0	0
*	**572 categorized queries** (82.30%)	110	362	100

Table 3.5 Genes related to immune, stress response, pathogenicity, cell toxicity

The left column provides the EST or cluster ID, followed by data base accession number, the best match in either or both BLASTX/BLASTP, gene annotation, species similarity and e-value.

EST or Cluster ID	Accession No.	Annotation	Organism	e-Value
Toxins and predicted toxins like genes				
EST_310/45	ABZ10815.1	Pathogen-related protein-like protein	*Malo kingi*	1e-131
EST_423	dbjBAA92257.3	Rhamnose binding lectin STL3	*Oncorhynchus mykiss*	2e-24
EST_906/Ps_642	refXP_001628983.	Dermatopondin-like gene	*N. vectensis*	8e-12
EST_332-1/2/EST_933	XP_001633958.1	Proteins members of the ShK toxic family	*N. vectensis*	8e-12
EST_269	EF636902	Toxin A precursor (CfTX-1).	*C. fleckeri*	2e-04
EST_205	XP_001626350.1	Predicted protein	*N. vectensis*	7e-94
Proteolytic enzymes and their inhibitors				
EST_63/EST_1162	gbAAO12213.1	Serine protease 1	*Aurelia aurita*	4e-37
TSP-2 -ShK				
EST_540 Cluster 598:	refYP_001931573.1	Proteinase inhibitor I4 serpin	*Sulfurihydrogenibium sp*	6e-26
Cluster_803		zinc carboxypeptidase		1e-36
EST_617 Cluster 325:	refXP_001177881.1	Tissue inhibitor of metalloproteinase TIMP	*S. purpuratus*	2e-09
Cluster_53	refXP_001623822.1	Serine carboxypeptidase 1 precursor protein	*N. vectensis*	1e-94
Contig_897	spQ66IL0	Chitinase domain-containing protein	*Xenopus tropicalis*	7e-09
EST_458 in	gbEDL32181.1	cwcv and kazal-like domains cystatin.	*Mus musculus*	3e-05
Cluster_614		Inhibitor of Cysteine proteinase		
Cluster_23	embCAD11970.1	Pacifastin-related serine protease inhibitor		5e-07
Related to immune system				
Cluster_72		Cyclophilin A		
Contig_408	refNP_035279.2	peptidylprolyl isomerase B Cyclophilin B	*Mus musculus*	5e-42
Cluster_457	refNP_033942.1	Caspase 8	*Mus musculus*	81e-19
Cluster_456	gbAAR22506.1	Caspase 3	*Cricetulus griseus*	
Contig_250	gbAAO15713.1	Allergen Pen m 2	*Penaeus monodon*	6e-16
Related to signalling pathways and transcription factors				
Cluster_878	Q9N658	Zinc finger protein sens (Protein senseless	*Drosophila melanogaster*	4e19
EST_317	gbAAS99630.1	Zinc finger transcription factor Belongs to the krueppel C2H2-type zinc-finger protein family	*Acropora millepora.*	2e-09
EST 81/ Cluster_56		bzip transcription factor		1e-32
EST_136	embCAA06314.1	PMP1 protein	*Podocoryne carnea*	9e-50
EST 33		Eukaryotic initiation factor 1		
EST_307 in Cluster 131	gbABC88377.1	Tolloid	*N. vectensis*	2e-37
EST_370 /Cluster_620	refXP_697494.3	Kielin like	*Danio rerio*	2e-11
Ps_616/624	gbAAO60428.1	Chordin-like protein	*Hydra magnipapillata*	0.004
Molecular chaperones				
Cluster_218	gbAAO38780.1	Heat shock protein 70 heat shock protein 70	*Chlamys farreri*	6e-94
Cluster_194		HSP90 TNF receptor-associated protein		
Cluster_477	refXP_802129.1	similar to 71 Kd heat shock cognate protein	*Strongylocentrotus purpuratus*	9e-09
Cluster_745	gbABK27326.1	Mortalin similar to Stress-70 protein	*Lytechinus variegates*	6e-19
Contig_862	refXP_001512830		*Ornithorhynchus anatinus*	1e-51
Cluster_611	refNP_001011351.1	Hsp40 homolog	*Bos Taurus*	3e-09
Cluster_568	dbjBAF49513.1	Cytochrome c oxidase copper chaperone	*Branchiostoma belcheri*	6e-19
Involved in vertebrates Disease				
Contig_858	refXP_001249608.1	Similar to cerebellar degeneration-related	*Bos Taurus*	
Contig_877	refXP_001095869.1	Tumor protein, translationally-controlled	*Macaca mulatta*	8e-25
Contig_961	gbEDL19863.1	TP53 regulated inhibitor of	*Mus musculus*	7e-11
Cluster_216	refNP_001075461.1	antigen p97 (melanoma associated)		2e-6
Relevant miscellaneous				
EST_244 Cluster_97	gbAAM77229.2	Insecticidal toxin A2	*Xenorhabdus nematophila*	6.7
EST_512		Mini collagen. Nematocyst wall	*Hydra magnipapillata*	0.022
Contig_921	refXP_001629180.1	Ankyrin repeats containing proteins	*N. vectensis*	3e-27
EST_301en Contig_6	refXP_001009831.1	Nucleoside diphosphate kinase family protein	*Danio rerio*	2e-60
EST_203	gbAAF33234.1	Transferrin	*Paralichthys olivaceus*	5e-19
Contig_250	gbACB46943.1	Arginine kinase	*Cancer irroratus*	59 1e-07
EST_981 Cluster_62	refXP_001253867.1	Similar to neuropeptide OPSIN (G protein-coupled receptor 1)		
EST_192	gbABH07379.1	Voltage-dependent anion channel VDAC-3	*Paralichthy.*	5e-52

3.4.5.1 Signaling pathways and Transcription factors

Many major signaling pathways are found in members of Cnidaria including Wnt, Hedgeog, RAS-MAK, Notch and TGFβ (Technau 2005). Four cDNA clones encoding molecules involved specifically in the TGFβ pathway were found in Cubozoa (see Table 3.6), including a bone morphogenesis protein (EST_136 BMP1) and three secreted antagonists, Tolloid, Cordin and Kielin-like. However, not all the molecules involved in the TGFβ cascade were detected.

3.4.5.2 Proteins involved in stress

The translated sequences of two *M. kingi* ESTs shared high homology with the two most abundant secreted proteins found in the nematocysts of other box jellyfish such as *C. fleckeri* CfTX-1 (43 kDa) and CfTX-2 (45 kDa) (Brinkman & Burnell 2007), *C. rastoni* CrTXs (43–45 kDa), *C. alata* CaTX-A (43 kDa) and *C. quadrigatus* CqTX-A (44 kDa) (Nagai 2003). These proteins have been reported to be lethal haemolytic proteins. The cDNA sequences identified in the *M. kingi* EST analyses were truncated cDNA clones. The finding of these sequences in the small *M. kingi* data set suggest strongly that these genes are highly expressed transcripts in adult Irukandji jellyfish and correlate well with the SDS-PAGE venom profile of other box jellyfish, in which apparent homologs of cytolytic proteins have been shown to be the major constituents.

A protein belonging to the stress response category (encoded by EST_906) showed homology with a novel dermatopontin-like cytolysin from the fire coral *Acropora sp.* (Iguchi *et al.* 2008). Interestingly, this protein comprises a six-cysteine domain resembling the ShK[1] toxic motif. This motif is of great relevance to this research because it has been found in a diverse array of cubozoan proteins with apparently different functions.

[1] ShK stands for a structurally defined toxic polypeptide from *Stichodactyla helianthus.* Later it was found that ShK is a complete encoding product of a gene. Homologs have been reported in other anemones: Metridin (*Metridinium senile*), BgK (*Budosoma granulifera*),and AeI (*Actinia equine*). All of these peptides are powerful inhibitors of T lymphocyte voltage-gated potassium channels (Kv1.3) Kath, J. C., Hanson, D. C. & Chandy, K. G. 1997 T lymphocyte potassium channel blockers. *Annual Reports in Medicinal Chemistry, Vol 32* **32**, 181-190.

3.4.5.3 Transcripts involved in pathogenicity.

We identified an outstanding EST encoding a secreted protein with a cysteine-rich domain similar to the ShK toxic peptide. This is named Mk phatogenin (protein ID Accession no ABZ10815), interestingly this jellyfish protein shares similarities to a subfamily of the plant pathogenesis-related protein (PR) superfamily. Pathogenesis-related proteins were first identified in plants and are so named because they are produced in elevated levels upon pathogen-induced injury or other stress (Van Loon & Van Strien 1999). PR proteins are not limited to plants and members have been reported in diverse phylogenetically unrelated species from bacteria, viruses and animals (Geer *et al.* 2002). The finding of these genes in ancient animal-like jellyfish is relevant and represents a link to elucidate the evolution of pathogenesis proteins. The mRNA sequence is under the accession number EU410456 .

3.4.5.4 M. kingi *proteases and inhibitors*

It is well established that marine cnidarians host a variety of proteases (Adhikari *et al.* 2007; Rojas & Doolittle 2002) and inhibitors (Chera *et al.* 2006), but none have been described in Cubozoans. Our study demonstrated that adult Irukandji box jellyfish abundantly express proteolytic-related transcripts e.g. elastase precursors, plasminogen-like and kazal-like domain containing proteins (Cluster_614), trypsin serine proteases -TSP- (Engel *et al.* 2001) and transcripts similar to a large family of metalloproteases. A serine protease and a metalloprotease inhibitor named box jellypin and metallopin, respectively, were also recognized.

Two different TSPs were identified, one of them containing an ShK toxic domain (EST_63, named Mk-TSP-2). Interestingly, similar protein architecture is found in the scyphozoan *A. aurita.* An analysis of the protein sequence places the jellyfish as a member of Group S1A (Rojas & Doolittle 2002), although the authors failed to mention that the C-terminus of this protein has a domain containing six cysteines with a high degree of similarity to the ShK toxic domain. This protein arrangement has not been found in sponges, bacteria, fungi or viruses and it is assumed here that this ShK-TSP domain combination arose in cnidarians. Furthermore, searches of StellaBase (*N.*

vectensis genomic database) detected homologs to this serine protease in anemone.

Serine proteases were found to be highly expressed in an EST analysis of the tentacles of the scyphozoan, *C. capillata* (Yang *et al.* 2003). Unfortunately, the sequences are not publicly available and no comparison could be made.

Proteolytic enzymes are present in all organisms, having diverse biological functions. In particular, serine proteases (SPs) are abundant peptidases in the animal kingdom and are present in viruses and bacteria but are rare in fungi. In venomous animals, SPs are often confined to vertebrate venom glands (Wagstaff & Harrison 2006) and Arthropoda salivary glands, and may be involved in pre-oral or even oral digestion (Batista *et al.* 2008). They may also interfere in the blood clotting and the platelet aggregation systems of prey, as seen in SPs from other animals with similar venom components. Some *in vitro* haemolytic activity of the anemone tentacle extract of *Paracondylactis indicus* suggests the presence of enzymatic components like phospholipases along with proteases, which may be contributing factors for the haemolytic activity of venom (Adhikari *et al.* 2007). In scyphozoans it has been reported that there is alpha chymotrypsin-like serine protease activity in the venom of *Rhopilema nomadica* (Gusmani *et al.* 1997). Additionally SP inhibitors can potentially inhibit bacterial and fungal proteases (Muta & Iwanaga 1996) as well as being implicated in self and non-self recognition.

Two true metalloproteases were present in the *M. kingi* survey. These were a bone morphogenic protein (PMP1) and tolloid-like astacins. Tolloid-like astacins were previously mentioned due to their involvement in the TGFβ signaling pathway (See Table 13). Astacin proteases are found in bacteria and are ubiquitous in the animal kingdom. They serve a variety of physiological functions such as digestion, hatching, peptide processing, morphogenesis and pattern formation (Mohrlen *et al.* 2003). Based on a phylogenetic analysis within eumetazoans, five clusters were revealed. The bone morphogenic protein and tolloid-like astacins are present in all eumetazoans, the mefrins are found only in cnidarians, two clusters are taxon-specific and the other cluster represents an astacin that probably evolved after the split of

cnidarians. The developmental role of astacins in cnidaria is not the same as the role astacins play in the higher animals (Mohrlen *et al.* 2006). A typical astacin comprises a presequence, prosequence and astacin domain, 0-3 ShK toxic motifs, a CUB domain (0-5) and up to two thrombospondin-1-like domains (TSP1). The astacin structural diversity seems to be due to gene duplications and potentially, at least in worms, all these genes could have different functions (Mohrlen *et al.* 2003). The toxic domain of these type of proteinases, at least in cnidarian members, could serve as a toxin to maintain prey paralysis during digestion (Pan *et al.* 1998).

3.4.5.5 Other protein members of the ShK toxic domain family (EST_332-1 and 2 and CbXI and CbXII)[2]

The ShK domain as mentioned above was found in a variety of *M. kingi* proteins identified in this research, e.g trypsin serine proteases, metalloproteases, a pathogen-related protein, a dermatopondin-like gene that is similar to a toxin from fire coral (Iguchi *et al.* 2008) and a group of basic preproproteins was identified to be composed exclusively of a tandem arrangment of the ShK domain and these derivates from EST_332 (Discussed in chapter 5

Interestingly, the complete cDNA derived from EST_332-1 has pairs of basic amino acids (lysine and arginine) located between some of the ShK toxic motifs strongly suggesting that proteolytic events at specific cleavage sites may be responsibe for the release of one or two peptides from their respective precursor molecules. In fact similar protein processing strategies are seen among sea anemone neurotoxins such as calitoxin from *Calliactis parasitica* (Spagnuolo *et al.* 1994) and other well documented prepropeptides with neurotoxic activity in cnidarians (Honma *et al.* 2005b; Ovchinnikova *et al.* 2006). Moreover a common motif ending in lysine and arginine in proparts in some cnidarian neurotoxic peptides has been previously reported (Anderluh *et al.* 2000b).

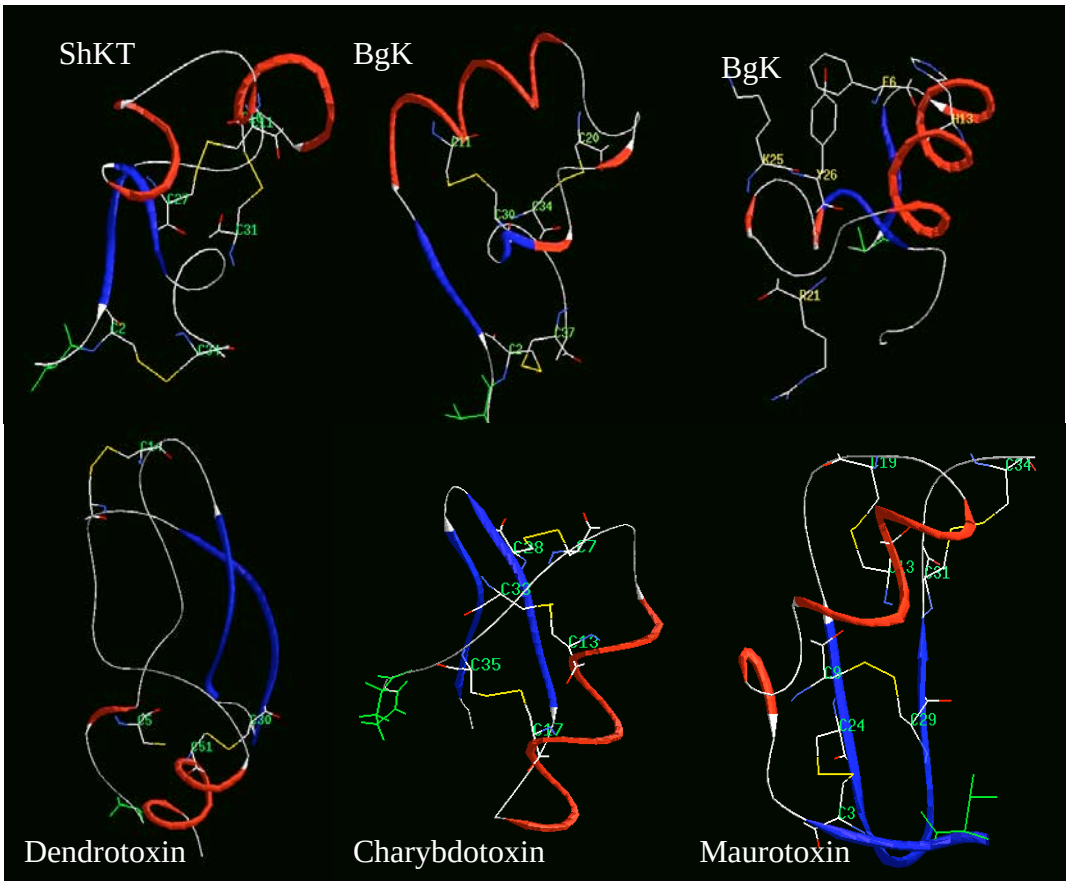
[2] These Cubozoa members of the ShK domain family group, their full-length cDNAs, conceptual proteins, gene expression patterns along with homologs detected in the recently release of *N. vectensis* will be discussed further in Chapter 5.

Searches of protein and conserved domain databases showed that the gene architecture of *M. kingi* EST_332 is similar in the genome of the starlet anemone *N. vectensis* and parasitic invertebrates such as *Caenorhadbitis elegans, Ciona intestinalis, Cryptosporidium parvum* and in the recently released *Brugia malayi* genome.

It is possible that a single ShK domain, or a combination of two or three ShK domains, may have the potential to act as a potassium channel blocker, which could only be determined by a functional assay. The similarities at the cDNA level (UTRs and ORFs) suggest that these genes arose from multiplication of this motif. The presence of ShK in other genes encoding functionally diverse proteins (Mk TSP-2, Mk pathogenin and Mk dermatopondin) may be due to gene insertion of ShK. The coding region of an ShK ortholog in anemones is intronless (Anderluh *et al.* 1995; Gendeh *et al.* 1997a), and this may be the reason by which this domain is easily transposable. Southern and northern blot analyses would assist in the elucidation of this matter. On the other hand, because the *M. kingi* transcriptome is not complete, it is probable that other members of this enigmatic ShK domain-containing protein family will be identified.

It is interesting that anthozoan peptides active in potassium channels show a low evolutionary rate in the six cysteine residues (see figure 3.6) in which it is particularly evident that the first three residues are invariably (**C**X**D**) and three out of the six cysteine residues are located within the the last eight residues at the the N-terminus with the signature **C**X**KTC**XX(X)**C**. This indicates the importance of cysteines in the formation of disulphide bridges for the maintenance of tertiary structure. Several authors have reported potassium toxins (Cotton *et al.* 1997; Minagawa *et al.* 1998) and there are key amino acids on these toxins, usually a functional amino acid dyad of lysine and an aromatic residue that may be linked to potassium channel binding. Intriguingly, the functional dyad is superimposed on the tertiary structure in toxin blockers of potassium channels in some invertebrates and vertebrates (Dauplais *et al.* 1997). In fact the small K-channel-blocking peptides in anemones are about the same size as scorpion K-channel toxins and have similar secondary structure (Kem *et al.* 1996). This feature is shown in the 3D models in figure 3.6 indicating the cysteine

residues and the functional dyad of several invertebrates and vertebrate toxins active in K^+ channels. Cubozoans ShK domains identified in this research also contain the same cysteine signature sequence and this and other features are discussed in Chapter 5.



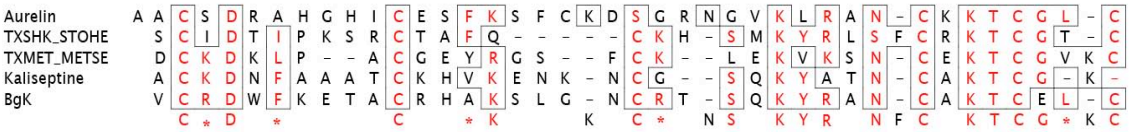


Figure 3.6. Six tertiary structures of peptides blockers of potassium channels

Peptides blockers of potassium channels adopt distinct scaffolds, however a functional amino acid dyad of lysine and an aromatic residue superimpose on the three dimensional structure of *Stichodactyla helianthus* ShK, *Budosoma granulifera* BgK, Snake Dendrotoxin, Scorpion Charybdotoxin and *Scorpio maurus* Maurotoxin. The amino acid sequence alignment at the bottom of the figure includes five cnidarian members of ShK-like toxins; conserved amino acids are shown in red.

3.4.5.6 Proteins involved in the immune system

Several potential immune-related transcripts were identified in the data. This included a transcript encoding ferritin, a major intracellular iron-binding protein in both prokaryotic and eukaryotic organisms. Ferritin plays a role in protecting cells from oxidative stress (Epsztejn *et al.* 1999) and is responsible for sequestering free iron from bacterial pathogens (Balla *et al.* 1992).

A transcript encoding cyclophilin A was also identified. Cyclophilins have diverse regulatory functions in mammalian cells, but they can also be involved in viral attachment to cells (Saphire *et al.* 1999) and in stress response to oxygen depletion (Santos *et al.* 2000). Transcripts encoding caspases 3 – 8 were also identified. Yeast cells lack caspases and therefore do not undergo apoptosis. It may be that this fundamental process originated in early metazoans (Wilkins 2000).

3.4.6 Non-metazoan genes in M. kingi

In our data for Cubozoa, a small percentage of genes were identified to be bacteria-like. Clones suspected to be contamination from marine bacteria were eliminated and true bacterial-like clones were identified by BLASTN sequence similarity searches. Sequences containing similarities to plant genes and other non-bacterial prokaryotic genes were also found, but returned low e-values (See Table 3.7).

The presence of prokaryotic, fungal and virus-like clones has been reported in *Acropora millepora* and *N. vectensis*. Non-metazoan-like genes are also represented in *Hydra* or *Hydractinia* ESTs. Furthermore, Southern blot analyses conducted for nine of the *Acropora* genes in this category confirmed their presence in the *Acropora* genome (Technau 2005). The same authors suggest that this small percentage of genes in cnidarians represent preserved ancient genes that were present in a common ancestor but have been lost in higher animals. Not only have prokaryotic, fungal and virus-like genes had been found in cnidarians, but also plant-related genes, proposed to be acquired by horizontal transfer. These were identified in hydrozoan species where the gene products are active. Examples of these include a phytohormone and an ascorbate peroxidase (Habetha & Bosch 2005; Puce *et al.* 2004). This is really quite notable because the preservation of active or otherwise dormant non-metazoan genes may confer cnidarians an advantage to cope with forthcoming environmental challenges.

Table 3.6 cDNA tags similar to non-metazoa genes

Bacterial-like clones (blue), yeast (orange), fungi (violet) and plants (green).

ID	DB identifier	Annotation	Organism	E.value
EST_165	EAY32189.1	Anaerobic dehydrogenases	*Vibrio cholerae*	3e-37
EST_330	ZP_01590040.1	Class II aldolase, tagatose bisphosphate	*Escherichia coli*	1 e-131
EST_560	AAC39773.1	Hepatitis A virus cellular receptor 1	*Theileria parva*	2e-04
Cluster_30	XP_001351961.1	hypothetical protein PFI0430c	*Plasmodium falciparum*	
Cluster_531	EAA17738.1	Hypothetical protein	*Plasmodium yoelii yoelii*	0.87
Cluster_536	XP_456667.1	Hypothetical protein DEHA0A07942g	*Debaryomyces hansenii*	2.9
Cluster_724	XP_001314870.1	Hypothetical protein TVAG_252980	*Trichomonas vaginalis*	6e-04
Cluster_555	ZP_01881133.1	Heme exporter protein CcmC	*Roseovarius sp. TM1035*	3.2
Cluster_584	EAW08579.1	Glutathione S-transferase, putative	*Aspergillus clavatus NRRL 1*	6.4
[3]Cluster_552	BAC44521.1	Predicted integral membrane protein..	*Mycoplasma penetrans*	0.85
[4]Cluster_755	CAH17867.1	Inositol phosphorylceramide synthase	*Pneumocystis carinii*	7e-25
Cluster_739	BAB40956.2	Myo-inositol-1-phosphate synthase	*Avena sativa*	0.047
Contig_840	CAN61340.1	Hypothetical protein	*Vitis Vinifera*	9e-46
Contig_1015	EAY97155.1	Hypothetical protein OsI_018388	*Oryza sativa*	3e-20
Cluster_588	EAZ08436.1	Hypothetical protein OsI_029668	*Oryza sativa*	9e-05

[3] Homologs in *N. vectencis* ref|XP_001638229.1| predicted protein

[4] *found homolog in cnidaria ref|XP_001631068.1| predicted protein N. vectens... 331 2e-89

3.4.7 ESTs have an inherent vulnerability for errors.

Overall, the sequence quality of individual EST sequences in this study was acceptable. Some sequences with high G/C content produced poor quality sequences under standard sequencing procedures. As ESTs are sequenced only once and represent only a portion of mRNA, this approach has a high susceptibility for error. In some cases base calling errors can occur, particularly for repeated bases such as G/C, although there were also complications with A and T repeat regions. The data were passed through stand-alone software and manually filtered to overcome possible ambiguities. Only six As of the poly A tail were included when a poly A tail was encountered.

Organizing and analyzing the *M. kingi* sequence data depended upon extensive use of internet-based strategies. Several of the tools used were not created to manage EST data sets. Although valuable direction was obtained, several gave misleading results and on several occasions hampered the systematic extraction of biological information from the EST collection.

A small proportion of the ESTs contained full-length sequences (70), while the vast majority of sequences were incomplete. Therefore the deduced protein sequences obtained from translated EST sequences were often not long enough to allow identification of protein motifs and domains. In addition, some ESTs consisted of long UTRs or non-coding mRNAs and these sequences yielded inaccurate predictions.

Low prediction accuracy in the analysis of secreted/non-secreted proteins was encountered and many dubious retrieved queries had to be manually identified. In contrast, better predictions were achieved during EST analysis by minimizing human interaction. However, after completing these analyses, some of the functional categories and gene ontology classification predictions seemed unrealistic. As it was a relatively small set of sequences, the ability to identify misleading interpretations was feasible. I believe that the algorithms used in the internet-based tools are designed for analysis of complete coding regions rather than fragmented products.

Also, reduction of analysis time is important because there is an enormous and continually growing amount of sequence data deposited in GenBank and molecular information has to be continuously updated in order to detect new homologies. A particular case in point was the discovery of MkTX-A and B in this study. These cDNAs were regularly searched during this PhD research but were not identified until the sequences of CfTX-1 and CfTX-2 (Brinkman & Burnell 2007) were deposited in GenBank, even though at that moment other related box jellyfish toxins were publicly available. The same applied with the homology found between Mk-dermatopondin and the novel cytotoxic peptide from fire coral (Iguchi *et al.* 2008).

Finally, perhaps the most important factor for the success of this EST analysis was the quality of the library. Often ESTs produced from non-normalized libraries need to be handled with care because they are subject to sampling bias resulting in under representation of rare transcripts. Subtraction and/or normalization measures have to be undertaken to produce a representative cDNA library (Bonaldo *et al.* 1996). This was not the case in this research and fortunately the *M. kingi* cDNA library proved to be non redundant, and the expressed genes sampled from the library were homogeneously represented (Figure 3.1). Although it is likely that in such a small data set the vast majority of tags represented highly abundant transcripts.

3.5 CONCLUSION

The sequencing of over 1,800 ESTs from a whole box jellyfish cDNA library has yielded new information and represents the first collection of ESTs for cubozoans. This project has also demonstrated the effectiveness of this approach for Irukandji jellyfish research.

Since the ESTs come from a non-normalized primary cDNA library (without amplification), the putative toxic and other defence-related proteins found in the data set may represent highly expressed transcripts, as only a small proportion of the

transcriptome would have been sequenced. Analysis of this small EST data set allowed us to recognise that the genetic pool of Cubozoa is more similar to mammalians than to other invertebrates such as flies and worms. Interestingly, this analysis revealed that Cubozoa more closely resemble Anthozoa than Hydroza, challenging long held views in cnidarian evolution.

It is clear however, that the cubozoan transcript data generated is undersized. Fortunately with sequencing technologies becoming more sophisticated, the cost and time involved in transcriptome research will decrease. Undoubtedly a more complete analysis of cubozoans will enable us to close a gap largely neglected in the cnidarian-based research community.

CHAPTER 4

Occurrence of peptidases and peptidase inhibitors in the *M. kingi* EST library with emphasis on the identification, molecular characterization, spatial gene expression and recombinant over-expression of a trypsin-like family of serine proteases S1A from *M. kingi* and *C. barnesi*

4. 1 ABSTRACT

Peptidases and their inhibitors occur abundantly in virtually all living organisms and engage in a range of vital processes that in part are responsible for biological success. Although previously found in cnidarians, there have been no reports of genes encoding peptidases or peptidase inhibitors from cubozoans. This chapter reports on the identification of genes encoding members of four of the five peptidase families found in metazoans and nine assumed peptidase inhibitors, representing only a small fraction of the cubozoan degradome[5] and the complement of protease inhibitor genes in Cubozoa.

Homologous cDNA clones encoding chymotrypsin-like serine proteases S1A (Cb-TSP-1 and Mk-TSP-1) were identified by immunoscreening *M. kingi* and *C. barnesi* cDNA libraries with *C. fleckeri* antivenom (see Chapter 2). In addition, a distinctive gene was found in the EST collection (Mk-ShK+TSP-2) that encoded a protein containing an N-terminal toxin homology motif followed by a complete trypsin serine protease domain (see Chapter 3).

Occurrence and distribution of cubozoan TSP-like genes were investigated by whole mount *in situ* hybridisation using riboprobes. The analyses revealed differences in expression patterns between the two genes. Mk-TSP-1 showed intense hybridization in the epithelial tissue of the gastric cavity, mouth lips and tentacular nematocytes,

[5] Repertoire of proteases expressed by an organism

indicating its involvement in both external and internal digestion of prey. In contrast, *in situ* hybridization using Mk-ShK+TSP-2 RNA probe, showed expression of the gene only in the batteries of tentacle nematocytes. This result and the presence of an N-terminal toxin motif suggested that Mk-ShK+TSP-2 may have a dual function and affect the nervous system of prey while assisting in internal digestion. Genome searches in representative organisms revealed that the Mk-ShK+TSP-2 domain arrangement originated with cnidarians. It is noteworthy that while Mk-ShK+TSP-2 was found in anthozoans, scyphozoans and cubozoans, no similar gene structure was found in hydrozoans, one of the most extensively studied classes in the Cnidaria phylum. Mk-ShK+TSP-2 was expressed in bacteria and was immunologically reactive towards *C. fleckeri* antivenom (CSL AV).

4.2 BACKGROUND

Peptidases and their inhibitors are distributed among all kingdoms and represent an extensively studied group of proteins. Proteases were initially identified as gastric proteolytic enzymes involved in the nonspecific degradation of dietary proteins (Puente *et al.* 2003), but they also play vital and specific roles in the control of cell signalling, cell behaviour during development, extracellular matrix remodelling, fertilization, immunity, protein processing, fibrinolysis, blood coagulation, tumour invasion and apoptosis (David *et al.* 2005; Puente *et al.* 2003 ; Seipp *et al.* 2006; Technau *et al.* 2003).

The most recent version of the MEROPS database contains information on more than 2,000 peptidases and nearly 400 inhibitors[6] classified according to families and clans (Rawlings *et al.* 2008). There are six classes of peptidases, classified according to the chemical groups responsible for catalysis of peptide bond hydrolysis: glutamate, aspartate, threonine, cysteine, serine and metallopeptidases (Bond & Beynon 1995), with cysteine, metallo- and serine peptidases the most common types.

[6] It is a large task to describe the peptidases and inhibitors due to the large number of families and immense variety of functions. I have no intention to deal with them all and only a general description is given. If interested you are directed to the MEROPS database. The literature section is regularly updated and includes several reviews. There is also a comparative genomics section and the draft genome of *Nematostella vectensis* is included in the database.

The importance of peptidases in living organisms is evidenced by the fact that 2 to 5 % of the genome encodes these proteins. The last decade has witnessed significant research interest in both peptidases and naturally occurring inhibitors because of their medical, agricultural and biotechnology relevance (Barrett 2004; Bartoli *et al.* 2006; Rawlings *et al.* 2002). For example, 14% of the 50 known and putative human peptidases are under investigation as drug targets (Rawlings *et al.* 2004; Rawlings *et al.* 2006), and high expectations for treatment exist for various inhibitors of proteases. This is due to the increase in knowledge about the significant contribution of cysteine, aspartic and serine peptidases associated with pathogen infections and many human disorders. A few specific examples include the aspartate protease inhibitor, renin, in hypertension, cathepsin D in metastasis of breast cancer, beta-secretase in Alzheimer's disease, plasmepsins in malaria, HIV-1 peptidase in acquired immune deficiency syndrome, and secreted aspartate peptidases in candidal infections (Dash *et al.* 2003). Some serine protease inhibitors play important roles in regulating the activity of host defence mechanisms. For example, elafin and the secretory leukocyte proteinase SLPI, exhibit potent antimicrobial activity (Thompson & Ohlsson 1986; Wiedow *et al.* 1990), while several cysteine inhibitors modulate host immune responses against pathogens and therefore are medically important (Gawlik & Poreba 2005).

4.2.1 Peptidases and peptidase inhibitor diversity

Recently, a bioinformatic analysis by Page and Di Cera (2008), delineated trends in peptidase diversity within a comprehensive set of complete genomes, including ancestral lineages such as the choanoflagellate *Monosiga brevicollis* and anemone *N. vectensis*. Three relevant considerations arose from this work:

i) Not all families of peptidases are equally abundant or represented, but a common proteolytic core of sixteen peptidase families exists in all kingdoms

ii) The basal degradome expanded in multicellular organisms through gains in intra- and extra-cellular activity, with an assortment of approximately 34 eukaryotic-specific peptidases; of these a small fraction is unique to higher metazoans

iii) Only ten recently evolved families are restricted to metazoans, excluding yeast. In conjunction, these considerations are indicative of a novel selectivity mechanism and enhanced spatial and temporal regulation of peptidases rather than new peptidases; gene inventions were enough to achieve the complex diversity of proteolysis in life. The diversity of peptidases in living organisms is summarized in Table 4.1.

Table 4. 1 Peptidase diversity in living organisms

The sixteen peptidase families identified so far in the genomes currently sequenced are highlighted in dark orange. Although several organisms seem to have lost one or more of these families, these are the peptidase families found in all living organisms. The peptidases shown in orange represent the thirty-four peptidase families widely distributed in eukaryotic organisms playing numerous roles in intra- and extra-cellular biology. The six families of peptidases involved in ubiquitin- and SUMO-mediated protein turnover are most notable given their vital role in protein homeostasis. The last block of proteins shaded in yellow correspond to ten peptidase families corresponding to peptidases in non-yeast metazoans. These peptidase families play a major role in extra-cellular processes. The most notable exception is the caspases - the executioners of cell death. Both the M12A and M12B peptidase families have diverse functions and expanded in higher metazoans. All data were taken and modified for clarity from Page & Di Cera (2008).

Family	Clan	Common Name	Biol. function	Family	Clan	Common Name	Biol. function
C26	PC	-glutamyl hydrolase	Folate metabolism	M41	MA	FtsH peptidase	protein processing
C44	PB	Amidophosphoribosyltransferase	Purine synthesis	S14	SK	Clp peptidase	Protein turnover
M1	MA	Aminopeptidase N	Protein turnover	S16	SJ	LonA peptidase	Protein turnover
M16B	ME	Pitrilysin Signal	Peptide processing	S1B	PA	HtrA peptidase	Protein turnover
M20A	MH	Glutamate carboxypeptidase	Protein turnover	S33	SC	Prolyl aminopeptidase	Turnover and
M22	MK	Hsp70/DnaK homolog	Chaperone	S54	ST	Rhomboid Membrane	protein processing
M24A	MG	Methionyl aminopeptidase	Protein turnover	S8A	SC	Subtilisin	Protein turnover
M24B	MG	Aminopeptidase P	Protein turnover	T3	PB-	Glutamyltranferase	Glutathione
A1A	AA	Pepsin	Protein turnover	M14B	MC	Carboxypeptidase E	Protein turnover
A22A	AD	Presenillin peptidase	Cell signalling	M16A	ME	Pitrilysin	Protein processing
A22B	ADI	MPAS	Protein processing	M16C	ME	Mitochondrial	Protein processing
C12	CA	Ubiquitin C-terminal hydrolase	Protein turnover	M18	MH	Aminopeptidase 1	Protein turnover
C13	CD	Legumain	Protein processing	M49	M-	Dipeptidyl-peptidase III	Protein turnover
C14B	CD	Metacaspase	Protein processing	M67A	MP	Poh1 peptidase	Protein turnover
C15	CF	Pyroglutamyl-peptidase	Bioactive peptide	M67C	MP	AMSH deubiquitinating	Protein turnover
C19	CA	Ubiquitin-specific peptidase	Protein turnover	M76	M-	Atp23 peptidase	Protein processing
C1A	CA	Papain	Protein turnover	M8	MA	Leishmanolysin	Protein processing
C1B	CA	Bleomycin hydrolase	Protein processing	S10	SC	Carboxypeptidase Y	Protein processing
C48	CES	ENP peptidase	Protein turnover	S26B	SF	Signalase	Protein processing
C50	CD	Separase	Cell division	S28	SC	Pro-Xaa	Protein processing
C54	CA	Autophagin	Autophagy	S59	SP	Nucleoporin	Nuclear pore
C65	CA	Otubain	Protein turnover	S8B	SB	Subtilisin	Protein turnover
C78	CA	Ubiquitin-fold modifier	Protein turnover	S9B	SC	Dipeptidyl-peptidase IV	Protein processing
M10A	MA	Matrix metallopeptidase	Extracellular	S9D	SC	Dipeptidyl-peptidase V	Protein processing
M14A	MC	Carboxypeptidase A1	Protein turnover	T1A	PB	Proteasome catalytic	Protein turnover
C14A	CD	Caspase	Apoptosis	M12B	MA	Adamalysin	Extracellular
C46	CH	Hedgehog	Development	M2	MA	Angiotensin-converting	Bioactive peptide
C64	CA	Cezanne peptidase	Protein turnover	M43B	MA	Pappalysin	Bioactive peptide
C67	CA	CylD protein	Protein turnover	M50A	MM	S2P peptidase	Signal peptide
M12A	MA	Astacin Extracellular	biology				

Similarly, to peptidases, peptidase inhibitors have been recorded in many different organisms, resulting in a large number of clans. Some peptidase inhibitors have arisen on different occasions during evolution and they mostly occur in eukaryotes. Only three families are known from Archaea, two of which (I4-serpins and I42-chagasin) are present in all kingdoms including viruses. Distinct inhibitors, however, are most numerous in eukaryotes and none of the prokaryote genomes studied by Rawlings and coworkers (2004), and included in the MEROPS database, contain more than six genes encoding members of the families of peptidase inhibitors so far recognized, whereas all of the eukaryotic genomes contain tens or hundreds of these (Rawlings *et al.* 2004).

4.2.2 Regulation of peptidase and peptidase inhibitor activity

The proteins that inhibit peptidases are of great importance in maintaining balance in living organisms. Proteins are continually synthesized and degraded and the concentration of proteins in normal conditions is determined by the balance between the rate of synthesis and the rate of degradation; the differences in these rates result in anomalies. Uncontrolled proteolysis activity can be the response to several stress factors, including starvation, heat shock, and chemical insults. Therefore proteolysis is regulated by several key mechanisms, by specific inhibitors, or by regulatory proteins. There are four distinct strategies for regulation (Bartoli *et al.* 2006; Rawlings *et al.* 2004):

i) Presence of signal transit peptides; the vast majority of peptidases are initially synthesised as inactive precursors with an N-terminal pro-sequence that has to undergo a processing event leading to peptidase activation.

ii) The target may be regulated in a way that proteolysis occurs when required to control the levels of important proteins such as structural, enzyme and regulatory proteins, some of which are thought to contain recognition signals either at the N- or C-terminus, that mark them for early degradation

iii) The ability of the protease to meet their substrates

iv) Proteases can be under regulation of endogenous protease inhibitors that prevent peptidase activity.

Despite their opposing activities, peptidase and peptidase inhibitor genes are intimately related and the number of genes in both these groups has expanded in eukaryotes almost equally, although there are crucial differences between them. Unlike peptidases many of the proteins that inhibit them contain multiple inhibitory domains, ranging from 2 to 15. Some of these domains can be as small as 14 amino acid residues while the whole protein can be comprised of hundreds of amino acids. The reactive sites in inhibitors are often not conserved, whereas the peptidase active sites are strictly conserved. Consequently predictions of functionality based on primary structure cannot be made with certainty. Finally, peptidase inhibitor regions have duplicated and reduplicated within genes, whereas there are few single polypeptides with multiple peptidase regions (Rawlings *et al.* 2004).

Different families of peptidases and their inhibitors are associated with biological pathways. While some of the specific activities can be evolutionarily conserved in some cases, functions of apparently homologous genes may not be exactly the same in higher and lower metazoans. A small subset of examples in point are:

i) Cysteine peptidases (caspases) deal with programmed cell death in vertebrates (David *et al.* 2005). This process also occurs in the first metazoan animal phyla and it has been revealed that cell death in hydra is almost indistinguishable from that in higher animals, however caspases are also indispensable for hydra metamorphosis (Bottger & Alexandrova 2007; Seipp *et al.* 2006).

ii) Different developmental functions of higher astacins in metalloproteases has been recorded from hydra and anemones in which astacin-like metalloproteases serve multiple functions; pattern formation, hatching, peptide processing (Mohrlen *et al.* 2003) as well as in morphogenesis and digestion (Pan *et al.* 1998; Sarras *et al.* 2002).

On the other hand, the role of serine peptidases in immune responses (Hargreaves & Medzhitov 2005; Imler & Hoffman 2000; Söderhäll *et al.* 1996), threonine peptidases in protein recycling and aspartate peptidases in viral infection (Page & Di Cera 2008) are all well established in some metazoans but little is known of those genes from ancestral lineages such as cnidarians. Because of the essential role of proteases and protease inhibitors and the absence of information on cubozoan proteases and protease inhibitors, a search for these two groups of proteins in the EST collection was performed and compared with the complete set of proteases and inhibitor families from the recently released genome of the sea anemone *N. vectensis* (Sullivan *et al.* 2006).

4.3 EXPERIMENTAL PROCEDURES

4.3.1 Cloning, transcript identification, expression constructs and sequence analysis

mRNA of *C. barnesi* and *M. kingi* specimens were used to construct two cDNA libraries in a λ Uni-ZAP® XR vector (Stratagene) and packaged using a Gigapack III Gold Cloning kit (Stratagene).

A truncated clone Mk-BJA-11, antigenic towards CSL *C. fleckeri* antivenom, was identified and selected for full-length sequencing due to the relevance of the protein encoded.

To search for peptidases and peptidase inhibitors, the EST collection was submitted to the peptidase database MEROPS using the batch scan option at http://merops.sanger.ac.uk/. Abundance and occurrence of cubozoan peptidases was compared with *N. vectensis* predicted peptidases.

4.3.2 Full-length completion of peptidases of interest by radiolabelling of DNA probes

Unless stated otherwise, all primers were synthesized by Sigma-Aldrich. Polymerase chain reactions (PCR) were performed in 50 µL volumes, following the manufacturer's instructions using standard thermal cycling profiles: 95°C (5 min), followed by 95°C for 1 min, 45 to 60°C for 1 min, 72ºC for 1.5 to 2 min, for 30 to 34 cycles, followed by an extension period of 72ºC for 5 to 8 min, then incubation at 4ºC. Samples were resolved on 1% (w/v) agarose–ethidium bromide gels and visualized by UV light.

The truncated cDNA Mk-BJA-11 clone was used as template to generate a 210 bp DNA fragment by PCR using gene specific primers: forward primer Gpr24 5'-TCACAGGATGGGGCAAGATG-3' and reverse primer Gpr25 5'-

TGGCGACAAACAAATGGTCC-3'. A DNA band of the expected size was excised from the agarose gel and purified (Qiagen). The purified DNA fragment (25 ng) was labeled with [α-^{32}P]dATP by random priming using a DECAprime II Labelling Kit (Ambion). The radioactive probe was used to hybridize approximately 8 x 10^4 plaques from each of the *C. barnesi* and *M. kingi* cDNA libraries. Positive clones were rescreened at lower plaque density and plasmid DNA isolated from single plaques by *in vivo*-excision (Stratagene). Resulting plasmids were sequenced in both directions (Macrogen, Korea).

Nucleotide and amino acid sequence analyses and comparisons were primarily performed using MacVector and Sequencher software packages. Guide phylogenies were constructed with either ClustalW or MacVector.

4.3.3 Design of recombinant expression constructs

The cloning strategy involved two steps; for the first round of amplification the primers in Table 4.2 were used. Amplicons of the expected sizes were initially ligated into pGEM-T^{TM} (Promega) and, following digestion with appropriate restriction enzymes, subcloned into pProEX HT (Life Technologies) or pQE (Qiagen) expression vectors, previously linearized with the same set of enzymes. Recombinant clones were sequenced from both ends to confirm their identities and to ensure cloning in the correct reading frame.

Table 4.2 primers used for amplification of constructs

Gene	Forward 5'- 3'	Reverse 5'- 3'
Cb-TSP-1	Gpr65 SphI: GCATGCAAAACTACGACAAACCTAAG	Gpr66 Hind III: GGTACCGTAACGAATGTATTTGTTTATC
Mk-TSP-1	Gpr40 Sal I: GTCGACGTATACTATCATTGCTTTTATCTTTG	Gpr41 Sal I: GTCGACGATAACTAATATATCTGTTCATCCAATC
Mk-TSP-2	Gpr63 SphI: GCATGCGGTCTTCGTACGTTGTTAG	Gpr64: KpnI GGTACCAATATTAATGTACTTGGAAATCC

4.3.4 Recombinant Mk-Shk-TSP-2 protein

E. coli (BL21 or NM522 competent cells) were transformed and single bacterial colonies were used to inoculate 5 mL LB/ampicillin (x4) and grown overnight at 37ºC. Each overnight culture was used to inoculate 2 L flasks containing 0.5 L LB/ampicillin (60 µg.mL^{-1}). Cultures were grown at 37ºC until an $OD_{600\ nm}$ between

0.5 and 1 was reached and protein expression induced with the addition of IPTG to a final concentration of 1 mM. Following addition of IPTG, cells were grown at 30ºC for 16 hrs. Bacterial cells were harvested by centrifugation, resuspended in 20 mL lysis buffer and bacteria mechanically disrupted by sonication (8 bursts of 20 sec of sonication with equal intervals of cooling in liquid nitrogen). Sonicated bacterial suspensions were centrifuged at 40,000 x *g* for 30 min, and then filtered twice through 0.22 micron filters. Both soluble and insoluble fractions were kept for further analysis.

Recombinant proteins were processed in two steps under non-denaturing conditions using Ni^{+}-Sepharose (GE Healthcare). Crude extracts were mixed with Ni^{+}-Sepharose beads for 2 hrs to allow binding to the nickel matrix at 4ºC. The slurry was packed into a 1 cm diameter column, washed once with PBS (137 mM sodium chloride; 27 mM potassium chloride; 10 mM disodium hydrogen phosphate; 2 mM potassium dihydrogen phosphate; pH 7.2) and 25 mM imidazole with an open flow of 0.2 $mL.min^{-1}$, and then washed with 4 volumes of column buffer containing PBS and 50 mM imidazole. After the last wash, the column was allowed to run dry, bound protein was eluted with 5 mL elution buffer (PBS and 250 mM imidazole) and 0.5 mL fractions collected. Insoluble, soluble and fraction samples were denatured by boiling for 5 min in sample buffer (0.125 M Tris-HCl, 4% (w/v) SDS, 4% (v/v) 2-mercaptoethanol) and analyzed. Protein samples were analyzed by both 10-15 % SDS-PAGE stained in Coomassie Blue R-250 and western blotting using nitrocellulose membranes and the Bio-Rad Mini-Protean II electrophoresis system. Proteins were detected on nitrocellulose membranes using mouse anti-tetra histidine antibodies (Qiagen) and rabbit anti-mouse antibodies conjugated to horseradish peroxidase (Sigma, St. Louis, Mo.).

4.3.5 Mk-TSP-1 and -2 whole mount mRNA **in situ** *hybridizations.*

The whole-mount hybridization method used was a variant of Hayward et al (2001), optimised for cubozoans. Briefly, the pBluescript SK plasmids containing the DNA insert were linearized using *Xba*I. Approximately 2 µg plasmid DNA was digested for 3 h at 37ºC and complete linearization was checked by running approximately 500 ng of DNA on a 1% agarose gel. Linearized plasmid was precipitated by addition of a

final concentration of 0.3 M sodium acetate (pH 5.2) and three volumes of ethanol, and incubated over night at -80ºC. DNA was recovered by centrifugation at 13,000 x *g* for 30 min, desalted by washing with 75% ethanol and re-centrifuged for 30 min.

Antisense RNA probes were synthesised using a digoxigenin (DIG)-labelling kit (Roche), following the manufacturer's instructions: reactions were assembled with 1 µg linearized DNA template, in a final concentration of 1X Transcription Buffer (50 mM Tris, pH 8.0, 8 mM magnesium chloride, 2 mM spermidine, 50 mM sodium chloride), 1X DIG-labelling mix (Roche) (3.5 mM DIG, 6.5 mM UTP, 10 mM ATP, 10 mM CTP, 10 mM GTP), 40 units ribonuclease inhibitor (Ambion), 74 mM DTT, 30 units T7 RNA Polymerase (Promega) and RNAse-free water (Ambion) to final reaction volume of (20 µL). Reactions were incubated at 37ºC for 3 h and terminated by adding 2 µL 25 mM EDTA pH 8; 2 µL of riboprobe was kept to check labelling efficiency. RNA probes were precipitated by adding 2.2 µL of sodium acetate 3 M, pH 5.2, (Clontech) and 50 µL of absolute ethanol. RNA was precipitated on dry ice for 2 hrs followed by 30 min centrifugation at 13,000 x *g*. The RNA pellet was resuspended in 50 µL of RNAse free water.

4.3.6 RNA hydrolysis

To regulate the length of the RNA, alkaline hydrolysis of the probes was conducted in a reaction of 55.5 µL containing 50 µL of resuspended RNA probe and 5.5 µL of sodium carbonate buffer (0.4 M sodium hydrogen carbonate, 0.6 M disodium carbonate pH 10.2). Incubation time of RNA hydrolysis at 60 ºC was calculated according to the equation

$$T = (L_o - L_f)/(k \times L_o \times L_f),$$

where T = time expressed in minutes, L_o = transcript length in kb, L_f = final length required in kb for a molecule to diffuse into tissues recommended to be 0.25 kb (Hayward *et al.* 2001), k = 0.11 (strand excision constant expressed in $kb.min^{-1}$).

Hydrolysis reactions for Mk-TSP-1 and -2 (1,100 and 1,200 bp, respectively) were incubated for 30 min and reactions were terminated by adding 2 µL 3 M sodium acetate pH 5.2, 20 µg of tRNA and three volumes of absolute ethanol. Precipitation took place at –80ºC for 24 hrs, followed by centrifugation at 13,000 x *g* for 15 min. Vacuum-semidried RNA probes were resuspended in 80 µL RNA probe resuspension buffer (50% TE Buffer, 50% formaldehyde, 0.1% Tween 20) and stored at -80 ºC until used. Otherwise 40 µl of each probe were dissolved in 1 ml of hybridization solution and added to samples three hours after samples were prehybridized in 4 ml of prehybridization solution for 3 hrs.

4.3.7 Preparation of cubozoan specimens for whole mount in situ *mRNA hybridization*

Juvenile and adult jellyfish were frozen in a liquid nitrogen bath and stored at -80ºC for 11 months prior to the experiment. Samples were thawed on ice. Once samples had regained their characteristic shape they were added to 2 mL of filtered sea water and formaldehyde (4% final concentration) and gently swirled. Samples were incubated at 4ºC for 24 hrs, rinsed twice for 5 min with 5 mL of PBS pH 7.2 and twice more with PBS. After the final rinse, the PBS buffer was replaced with RIPA solution (150 mM sodium chloride; 1% (v/v) Nodinet P-40; 0.5% sodium deoxycholate; 0.1 (w/v) SDS; 1 mM EDTA; 50 mM Tris, pH 8.0) and incubated for 24 hrs at 4ºC. Water was removed from samples by gradually increasing the concentration of ethanol in PBT (PBS; 0.1% (v/v) Tween 20), (50, 70, 90 and 100%). Samples were cleared by placing the jellyfish in 50:50 (v:v) ethanol-xylene for 10 min then for 4 hrs in pure xylene or until samples were completely transparent. Adult jellyfish were difficult to clear, particularly those with mature gonads, in which case samples were placed in warm xylene with rocking at room temperature.

Re-hydration of samples was achieved by sequential rinses of 5 min each with 50:50 xylene/ethanol (Heralde *et al.* 2008), three times with absolute ethanol followed by 25:75 and 50:50 PBS/ethanol, three times with PBT, once with 50:50 PBS/hybridization solution (50% formamide; 4x SSC; 1x Denhardt's solution; 50

μg.mL^{-1} heparin; 5% (w/v) dextran sulphate; 250 μg.mL^{-1} tRNA; 500 μg.mL^{-1} denatured sonicated salmon sperm DNA; 0.1% (v/v) Tween 20), two rinses with hybridization solution and the last wash incubated for 10 min at room temperature. The final wash solution was then replaced with 4 mL of previously warmed hybridization solution. Samples were incubated at 56.5ºC for 3 hrs. RNA probe (40 μL) was mixed with 1 mL hybridization solution and added to samples. RNA hybridization was conducted for 48 hrs at 56.5 ºC.

After hybridization, samples were rinsed once for 24-48 hrs at 56.5ºC then three times for 20 min each with warm *in situ* washing solution [4x SSC; 50% (v/v) formamide; 0.1% (v/v) Tween 20]. Samples were washed twice more for 20 min each with *in situ* washing solution at room temperature, then rinsed for 15 minutes with 50:50 (v/v) *in situ* wash solution/PBT, followed by six rinses of PBT of approximately 15 min each.

Samples were incubated with rotation for 2 hrs in a 1:1600 dilution of anti-digoxigenin-AP (αDIG-AP) in PBT solution. After a 30 min wash with PBT and replacement of the buffer, samples were left at 4ºC overnight to remove unbound antibodies and to minimise background labelling. Jellyfish were washed six additional times for 30 min each. Following two washes in NTMT (100 mM sodium chloride; 5 mM magnesium chloride; 0.1% (v/v) Tween 20; 100 mM Tris, pH 9.5) solution, samples were equilibrated in alkaline phosphatase buffer for 15 min then immersed in BM Purple (Roche) or individual NBT/BCIP reagents (Promega) diluted in alkaline phosphatase buffer. Samples were incubated with gentle agitation in darkness. Colour development varied from 10 min to 24 h and was terminated by extensive washing of the samples in PBT. To enhance clearness of jellyfish, samples were placed in 70% glycerol in PBT and stored at 4ºC overnight. Jellyfish were examined using a stereoscopic microscope (Leica MZ FLIII) and photographed using a SPOT digital camera and its corresponding PC software. Captured images were edited as required, using ArcSoft PhotoStudio® and Adobe® Illustrator CS ® for Mac OS X.

4.4 RESULTS AND DISCUSSIONS

The cubozoan EST data set was searched using BLASTP and the MEROPS database, resulting in the identification of 37 clusters, eight of which are putatively involved in inhibition of proteolysis (Table 4.3). The remaining clusters were identified as proteolytic proteins according to the MEROPS peptidase database search and are listed in Table 4.4. indicating classification of family and class name, and including the possible active site residues and the predicted metal ligands in the case of metalloproteases (Rawlings *et al.* 2006; Rawlings *et al.* 2008). The table indicates the top BLASTX result against the MEROPS database in correlation with the *M. kingi* ESTs.

The peptidases identified belong in four of the five existing families: serine and metallo proteases (S1A, M12 being the most abundant), followed by cysteine and threonine peptidases; but no aspartate peptidases were identified. Because the group of genes identified represent a small fraction of the cubozoan degradome and the complement of protease inhibitor genes, the results were compared with a far more complete analyses, in this case the complete degradome of *N. vectensis*. The recent release of the genome of the sea anemone *N. vectensis* (Putnam *et al.* 2007) has enabled the identification of a set of 474 genes coding for peptidases (Page & Di Cera 2008), providing a useful framework that reflects what can be considered the degradome of lower metazoans. In *N. vectensis*, as in other metazoans, there is bias in the gene abundance of members of serine, metallo, cysteine and to a lesser extent aspartate and threonine proteases. In Figure 4.1 the result of this analyses and the prevalence of cubozoan peptidases is shown with blue arrows. These results indicate that similar trends of occurrence are found in cubozoans when compared with *N. vectensis.*

Table 4.3 Peptidase inhibitors identified in the *M. kingi* EST collection

The gene name is based on gene similarity analyses. In column two, the peptidase inhibitor family is indicated followed by the cDNA region in column three. The column corresponding to BLASTX or P is given the best hit out of ten; indicating the database identifiers, gene product species similarity and the e.value.

Cluster ID	Gene name	Family	cDNA region	Merops ID hit	Region	E. value
MK_540	Box Jellypin	I04	119-451	MER023853	7-119	2.80e-12
MK_617	Box Metallopin	I35	441-722	MER018458	102-195	2.60e-09
MK_389/455	Pacifastin serine protease inhibitor	I19	286-369	MER030026	24-51	9.50e-08
Mk_620		I19	80-101	MER023615	32-53	1.20e-05
MK_231		I39	355-777	MER018486	861-1003	2.20e-30
MK_458	Inhibitor of Cysteine	I31	109-222	MER110968	237-276	4.60e-07
MK_749		I63	71-202	MER104414	770-819	7.60e-08

Table 4.4 Occurrence of peptidases and inhibitors within the cubozoan EST collection

Family/class name, cDNA region, predicted active site residues, predicted metal ligands (for metalloproteases), MEROPS ID, hit start-hit end and e-values were obtained from searches of the MEROPS database (Rawlings *et al.* 2006; Rawlings *et al.* 2008). Members of the same peptidase family are shaded with the same colour. The top BLASTX hits of the M. kingi and C. barnesi ESTs against the MEROPS database are also shown.

Cluster ID	Family/Class	Begin-end	Active site	Metal ligands	Merops ID	Hit begin/end	E. value	Best hit by BlastX
Mk_212	**S01A**	108-818	H148, D196, S294		MER001503	34-263	2.8e-52	ref|NP_446461.1| Chymotrypsin-like
Mk_101	**S01A**	222-932	H262, D309, S408		MER005656	32-258	7.9e-48	ref|XP_001623008.1| Chymotrypsin-like protease
Mk_770	**S01A**	2-556	H, D38, S136		MER002738	84-267	1.2e-33	(Q9UNI1) Elastase-1 precursor
Mk_951	**S01A**	622-747	H, D632, S		MER078076	107-150	2.e-5	gb|AAO12213.1| Serine protease 1 and 2
Mk_982	**S01A**	24-284	H, D, S85		MER000144	159-240	1.5e-12	(P08419) Elastase-2A precursor
Mk_578	**S01A**	2-646	H27, D75, S170		MER005926	232-446	5.4e-47	(O35453) Serine protease hepsin
Mk_521	**S01A**	79-297	H, D, S101		MER002683	176-242	6.9e-15	(P80646) Chymotrypsin B
Mk_53	**S10**	3-974	S43, D248, H309		MER118171	114-440	3.4e-78	refXP_001623822.1 Serine carboxypeptidase 1
Mk_216	**S60**	25-348	N87K, E279S		MER033291	24-355	1.2e-65	gb|EDO32451.1| Predicted protein
Mk_185	**S08A**	405-506	D, H, N, S		MER004988	345-378	3.3e-11	gb|ABC86483.1| Membrane-bound transcription factor
Mk_567	**S09X**	17-343	V73S, D, H		MER101644	1050-1151	6.4e-6	ref|XP_624088.1| Actin-interacting protein
Mk_1027	**S09X**	1-597	C55S, V154D, H		MER115415	943-1134	5.3e-28	
Mk_386	**S09X**	82-570	E99S, R194D, G240H		MER115419	636-785	1.1e-15	ref|NP_001015781.1| MGC108352
Mk_298	**S33**	75-194	S, D, H		MER033856	114-153	6.6e-6	gb|AAL02400.1| Dihydrolipoamide S-acetyltransferase
Mk_828		35-67	S, D, H		MER033261	45-77	4.e-5	
Mk_257	**S49**	377-682	S		MER001303	206-307	6.8e-55	gb|AAM88904.1| Guanine nucleotide-
Mk_131	**M12A**	214-789	E307	H306, H310, H316	MER118357	3-185	6.3e-5	gb|ABC88377.1| Tolloid
Mk_803	**M14A**	358-705	R, E	H423, E426, H	MER005813	121-238	3.e-28	(Q5U901) Carboxypeptidase
Mk_105	**M12A**	118-606	E181	H180, H184, H190	MER005122	80-242	6.3e-54	emb|CAA06314.1| PMP1 protein
Mk_106		7-161	S34H	H, D, L36H	MER121033	195-352	1.3e-5	
Mk_75	**M23B**	491-814	H	H, D, H	MER065278	2220-2327	7.3e-9	gb|AAZ40189.1| Non muscle myosin II
Mk_3	**M23B**	234-704	L300H	H, D, Q301H	MER050288	921-1077	7.9e-7	
Mk_1012	**M23B**	90-425	H189	H99, D103, H191	MER085783	98-206	1.9e-6	ref|NP_001041520.1| Leukocyte cell-derived chemotaxin 2
Mk_382	**M23B**	10-417	L102H	H, S10D, Q103H	MER050288	896-1031	6.e-14	Chromosome segregation ATPase-like
Mk_595	**M67X**	3-587	E	E11H, V13H, N19D	MER030133	160-354	2.2e-5	gb|AAY66837.1| Eukaryotic translation initiation fact. 3
Mk_750	**M67X**	5-286	E	H, H, D	MER021886	179-272	1.9e-21	ref|NP_001026122.1| Eukaryotic translation initiation fact. 3
Mk_456	**C14A**	238-525	H, C248		MER000853	153-247	1.8e-19	gb|AAR22506.1| Caspase 3
Mk_457	**C14A**	361-831	H480, C		MER002849	193-354	7.3e-16	No high similarity
Mk_927	**T01A**	2-274	T		MER000551	140-230	1.4e-32	gb|AAZ38994.1| Proteasome-like protein a venom gland-

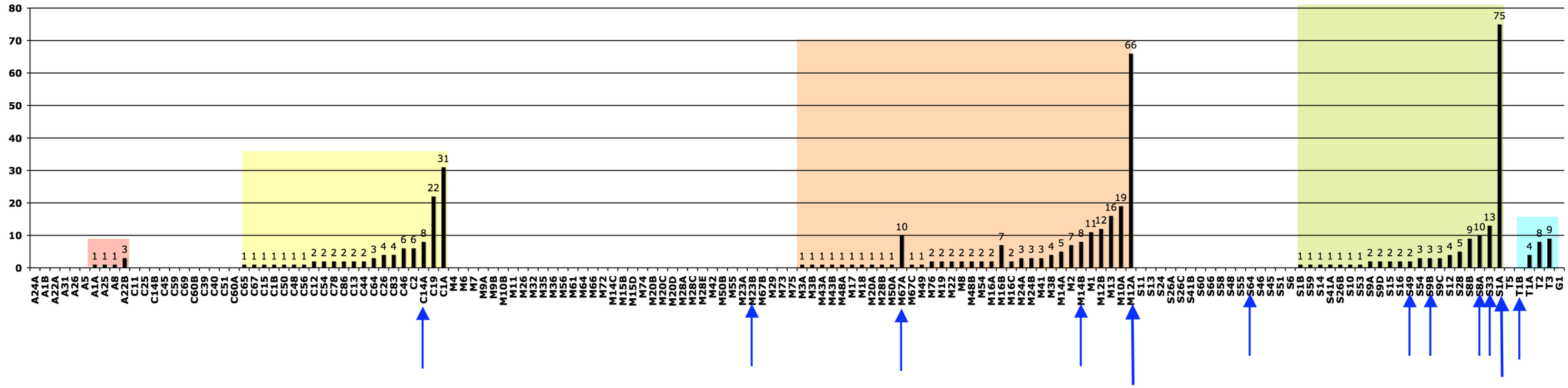


Figure 4. 1 Peptidase diversity within the cubozoan EST collection compared with the peptidase diversity of the starlet anemone *N. vectensis* draft genome

From the data gathered by Page & Di Cera (2008) 474 genes were found encoding proteases from the *N. vectensis* genome. Of the six peptidase families, only the small glutamate peptidase family (G1) is absent in the starlet anemone. Coloured shading corresponds to different peptidase families; red: aspartate; yellow: cysteine; orange: metallo; green:serine and blue: threonine.. Blue arrows indicate the assigned peptidase families/clans of the cubozoan ESTs obtained in this study. The Cubozoa EST collection contained members in four of the five families identified so far in cnidarians.

4.4.2 Identification of chymotrypsin-like serine peptidases: Cb TSP-1 and Mk TSP-1 and -2

Not all of the proteases found in the EST analyses were fully characterized in this study, and instead focus was directed towards those clones that were considered to be interesting, such as the peptidases: Cb-TSP-1 and Mk-TSP-1 and 2.

Chymotrypsin-like serine proteases, in particular Mk-TSP-1, attracted research attention because a truncated cDNA clone (Mk-BJA-11) encoding a fragmented protein named Mk-chymotrypsin 1-like serine protease was originally identified by immunological cross-reactivity towards *C. fleckeri* antivenom (Refer to Chapter 2). Because Mk-CSL AV-11 was not a complete encoding product it was concluded to be antigenic due to bacterial or phage protein cross reactivity with CSL box jellyfish antivenom. However, it was considered relevant because serine protease activity is commonly found in venoms of vertebrates (Hana *et al.* 2008; Shuqing *et al.* 2008). Protease activity has also been detected in venoms of medically-relevant cnidarians such as hydrozoans (Tamkun & Hessinger 1981), scyphozoans jellyfish (Calton & Burnett 1982) and even tentacular extracts of anthozoans display significant proteolytic activity (Adhikari *et al.* 2007). It has been proposed that protease transcripts may originate from the cells surrounding the nematocytes, or from the intracapsular content of the nematocysts (Calton & Burnett 1982). Other research workers confirmed the presence of ∝-chymotrypsin-serine protease activity in the venom of the scyphozoan jellyfish *Rhopilema nomadica* (Gusmani *et al.* 1997), although the gene was not elucidated. Recently the occurrence of chymotrypsin serine proteases in the scyphomedusa *Aurelia aurita* has been reported (Rojas & Doolittle 2002). Furthermore, in an EST analysis, alkaline proteases were identified as part of the basic active components identified in the venom of the scyphomedusa *C. capillata* (Yang *et al.* 2003b).

Given that serine proteases are an important group of peptidases, but have not previously been identified in cubozoans, full length Cb-TSP-1 and Mk-TSP-1 cDNAs were successfully cloned and sequenced in this study for further characterization (Figures 4.2 and 4.3 respectively). Comparisons of both nucleotide and protein sequences showed that both proteins are homologs. In addition, a transcript encoding a peculiar trypsin-like serine

protease, containing a six cysteine C-terminus domain with high homology to peptidic toxins from anemones was isolated (Figure 4.5).

4.4.3 Molecular characterization of Cb-TSP-1, Mk-TSP-1

The Cb-TSP-1 cDNA clone lacks the 5'-end including the initiating methionine but contains an adenylation signal as well as a poly A tail (Figure 4.3). The clone encodes a protein precursor of 275 amino acids. The conversion of the precursor to a mature polypeptide of 250 amino acid residues requires excision of the 24 residue signal peptide. The mature protein has a predicted molecular weight of 27.79 kDa and an estimated isoelectric point of 10.37. Of the amino acid sequence, residues 13 to 35 correspond to a transmembrane domain and residues 37 to 268 correspond to the trypsin serine protease domain including the catalytic triad.

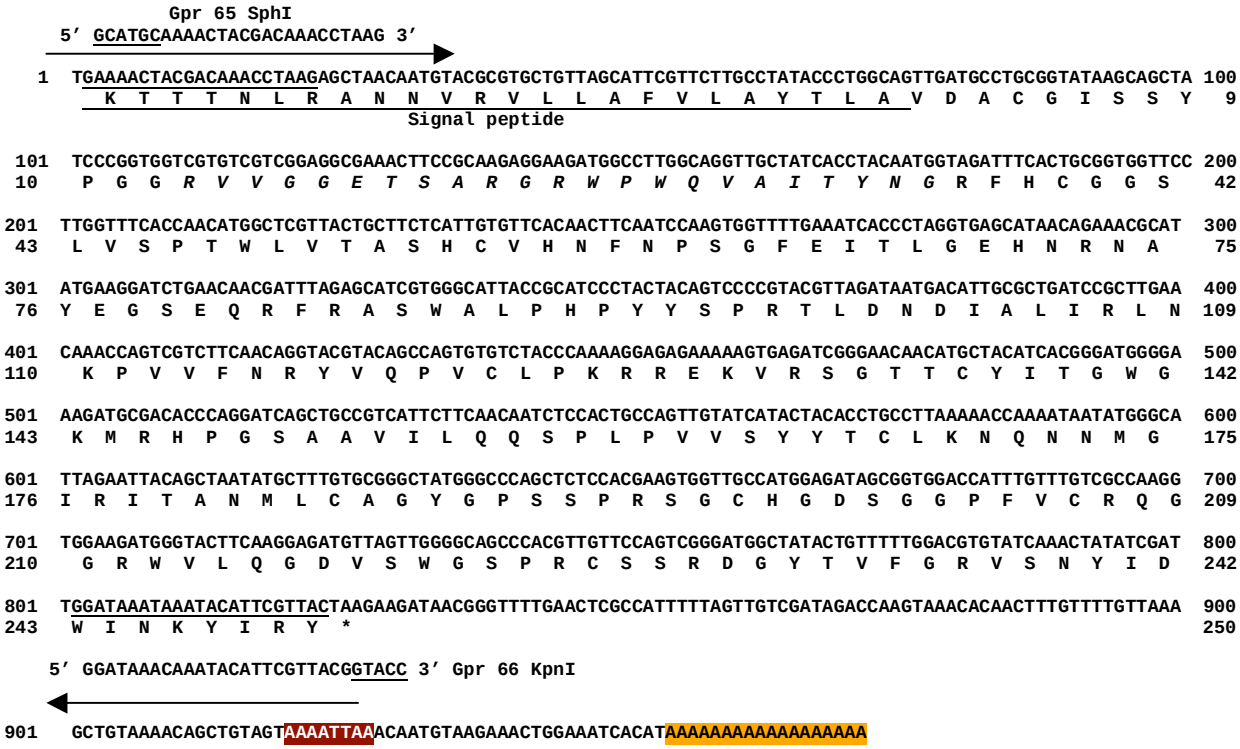


Figure 4.2 *C. barnesi* (Cb-TSP-1) cDNA clone (Allele 1) and amino acid sequence

Primers for expression clone construction are indicated on top of their respective sequence. Arrows indicate primer orientation.

Mk-TSP-1 (allele 1) is a cDNA clone of 1041 bp containing an initiating methionine an adenylation signal as well as a poly A tail (Figure 4.3). The clone encodes a precursor of 264 amino acids, 13 of which correspond to a secretion signal peptide. The mature protein of 251 amino acids has a calculated molecular weight of 27.7 kDa and an estimated isoelectric point of 10.11. Similar to its counterpart in *C. barnesi*, Mk-TSP-1 included the amino acid triad of an active TSP.

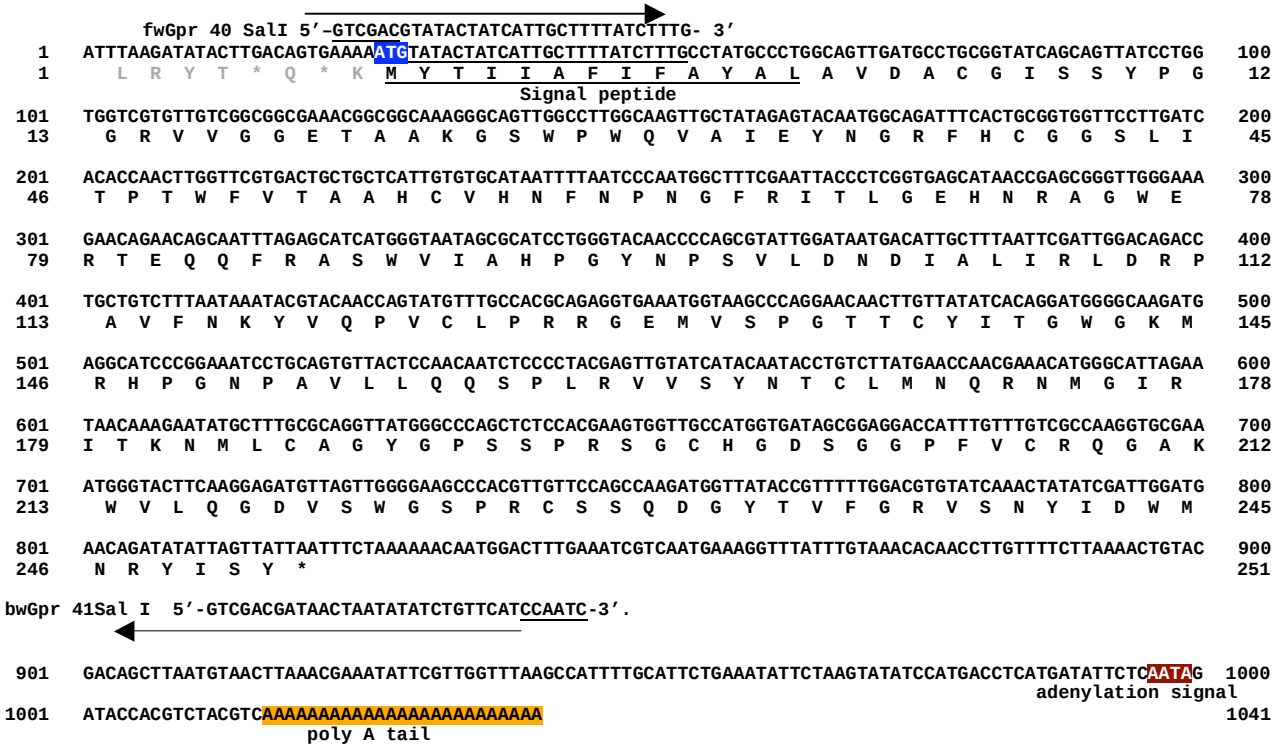


Figure 4.3 Mk-chymotrypsin-1-like (Mk-TSP-1) cDNA clone nucleotide and deduced amino acid sequence

4.4.4 Allelic variation in C. barnesi and M. kingi chymotrypsin–like serine proteases -1

The TSP-1 transcripts found in both *C. barnesi* and *M. kingi* cDNA libraries, show allelic variation. In the case of *C. barnesi* TSP-1, there were only two nucleotide substitutions at positions 411 and 870: G/A, A/G, respectively. At the protein level only the first nucleotide substitution affected the amino acid, at position 137: I/V. Both isoleucine and valine are aliphatic and hydrophobic amino acids. The major variation of Mk-TSP-1 at the protein level was 4 amino acid substitutions; no deletions or additions were observed.

4.4.5 Molecular characterization of Mk-TSP-2

The Mk cDNA library contains four alleles of Mk-ShK+TSP-2 with only four changes in the nucleotide sequence and these did not cause changes to amino acid residues. The longest Mk-TSP-2 clone was EST_63 (Figure 4.4). This clone lacked the 5'-UTR, but contained an initiating methionine, an adenylation signal and a poly A tail. The mature protein is 288 residues long and has an estimated isoelectric point of 11.8 and a predicted molecular weight of 33.05 kDa. The trypsin serine protease domain is indicated by the amino acids coloured in green. Mk-ShK+TSP-2 is a peculiar peptidase due to an N terminal fragment with a SXC/ShK toxin homology domain (highlighted in red, Figure 4.4) that is entirely absent in Mk-TSP-1 or Cb-TSP-1. This six cysteine motif is characteristic of metridinin-like toxins but SXC/ShK is unusual in chymotrypsin serine proteases. Comparative analyses indicated that the ShK+TSP domain combination may be an invention of cnidarians.

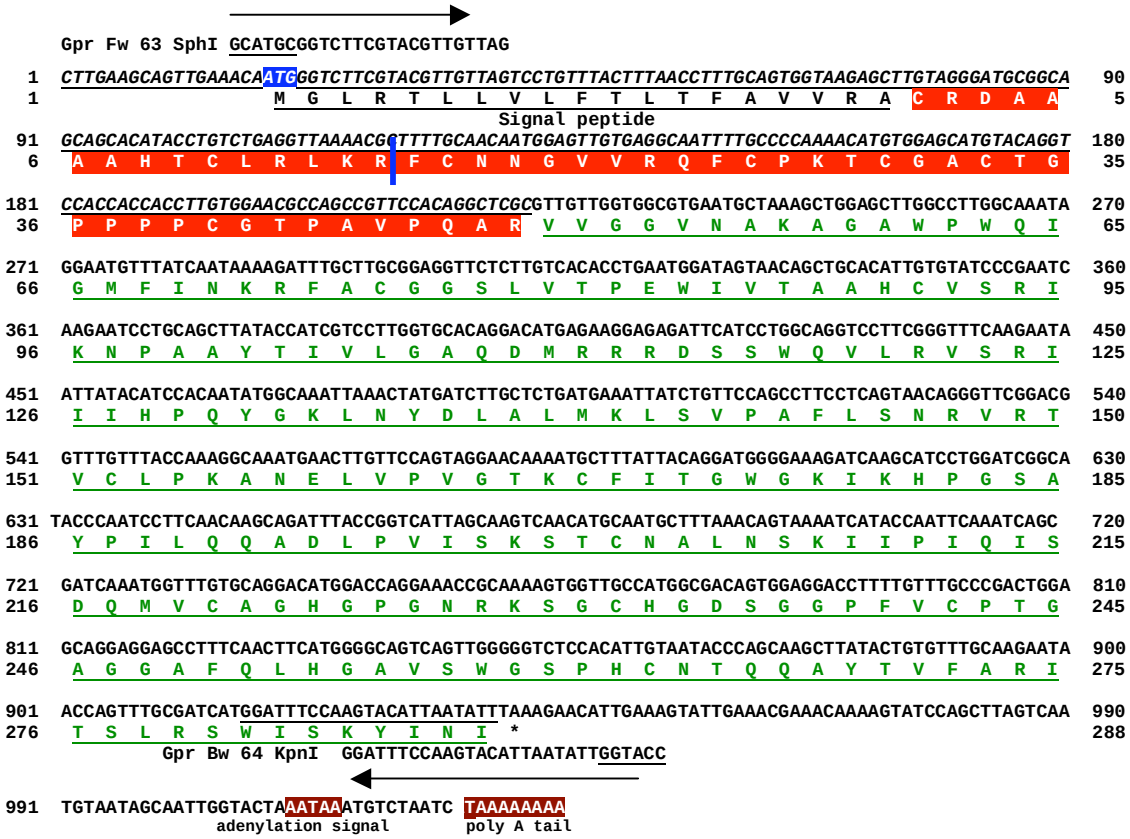


Table 4.4 Mk-chymotrypsin-2-like (Mk-TSP-2) cDNA clone nucleotide and deduced amino acid sequence.

The primers used for cloning the protein into the pQE30 expression vector are underlined in black omitting the restriction sites *Sph*I at the 5'-end and *Kpn*I at the 3'-end. Arrows indicate the direction of amplification.

The peptidase S1 family, clan A (chymotrypsin-like serine proteases) are not only highly expressed genes but they are also well conserved proteins in terms of secretion pattern, protein structure and important functional residues. For example, Figure 4.5A shows an alignment of TSP signal peptide sequences that is characterized by a traditional aromatic region, flanked by a hydrophobic region (Nielsen & Krogh 1998). Figure 5.5B illustrates (green lines) the region of Cb-TSP-1, Mk-TSP-1 and Mk-ShK-TSP-2 that contains the conserved trypsin serine protease domain (E-value=3.04e-83). The three cubozoan sequences in the alignment contain the residues that form the catalytic triad (H, D and S) in all metazoan TSPs. In the same alignment asterisks mark identical residues and dots indicate very similar side chains among the Cubozoa. The most highly conserved residues across TSPs in metazoans are boxed in blue. Notably, most of the cysteine residues and block of conserved residues flank the functional residues (Due to space limitations full alignment of sequences is not shown). If interested, a BLASTP search will generate a guide tree indicating the diversity and complexity of peptidases S1A among animals). Importantly, these non-mammalian serine protease genes (Cb-TSP-1, Mk-TSP-1 and -2) have 47-51 % sequence identity with mammalian genes, suggesting a common ancestral protease. Mk-TSP-2 is an unusual TSP due to the presence of a cysteine rich motif, SXC-tox 1/ShK (SMART DB; E-value=7.75e-81 (Schultz *et al.* 1998)), indicated with a red line in Fig 4.5 B. In addition MkTSP-2 has a furin-type cleavage site KR (threshold=0.5, Score= 0.542; ProP v.1.0b [Duckert *et al.* 2004)] in the SXC/ShK motif. This cleavage recognition site will be discussed further in the next section.

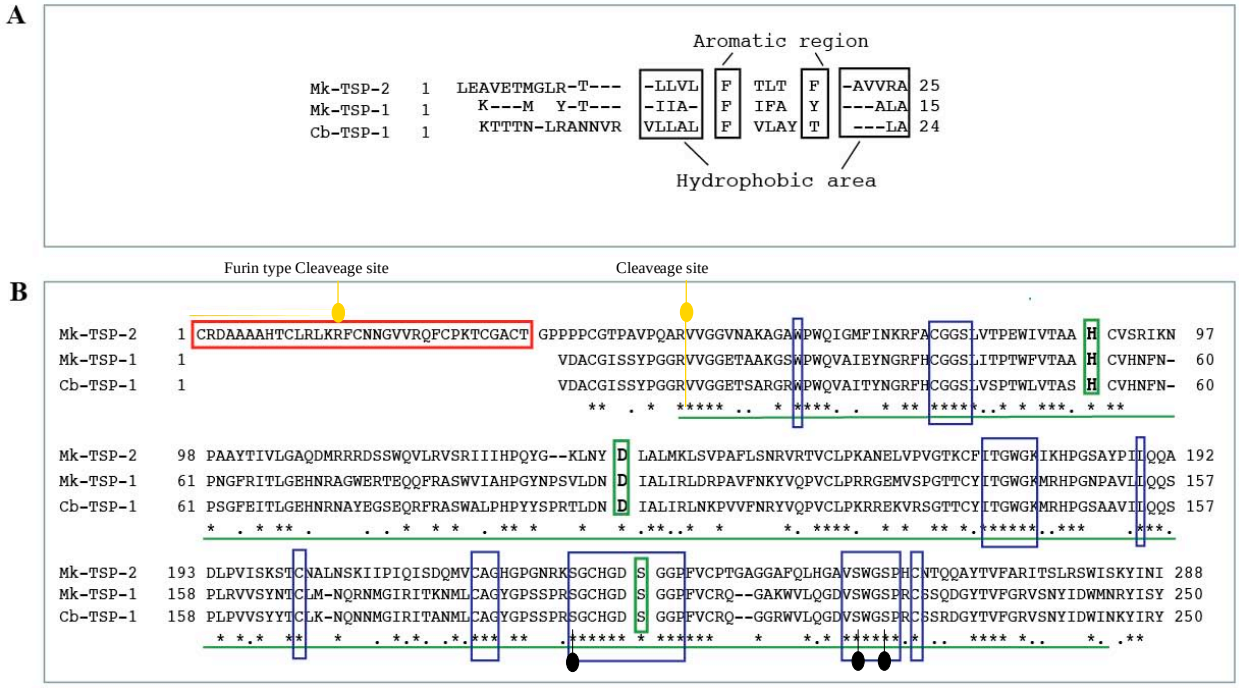


Figure 4. 5 Comparison of *C. barnesi* and *M. kingi* chymotrypsin-like serine proteases.

The yellow line and arrows indicates a propeptide region and cleavage sites, respectively. The ShK motif is boxed in red. The TSP domain is delimited in green, indicating with green boxes the catalytic residues (H, D and S**)**. Substrate binding site is marked with black arrows. Blue boxes indicate highly conserved residues among metazoans chymotrypsin serine proteases.

4.4.6 Expression constructs and recombinant protein expression studies of cubozoan peptidases Cb-TSP-1 and Mk-TSP-1 and -2

In order to study biological structure and function, significant quantities of proteins are required. However, since Irukandji jellyfish are difficult to obtain and yield low quantities of native proteins, an aim of this study was to produce recombinant protein for further analyses and use in other research laboratories.

In the case of peptidases Cb-TSP-1 and Mk TSP-1, the regions to be cloned were chosen to avoid unwanted proteolysis in the host bacteria. Thus, the precursors and not the mature encoding region (see Fig 4.2 and 4.3 for regions cloned into expression vectors) were cloned in the correct reading frame into pQE and pPROEX expression vectors. Correct insertion of

the coding region into the expression vectors was confirmed by sequencing. Despite the use of two different expression vectors, two different *E. coli* strains (NM522 and BL21), expression at several temperatures and different induction periods, neither Cb-TSP-1 nor Mk TSP-1 was expressed. SDS-PAGE and western blot analyses demonstrated that Cb-TSP-1 and Mk TSP-1 could not be expressed in bacteria. Consequently, we were unable to confirm whether a complete and functional cubozoan TSP-1 cross reacts with CSL box jellyfish antivenom. In spite of the outcome, a different approach was taken to shed light on the probable function of the product of this gene, as discussed further in Section 4.4.7.

The protein sequence of Mk-TSP-2 was back-translated to obtain DNA sequence and an expression construct was designed to contain a 5'-*Sph*I and a 3'-*Kpn*I site to clone into pQE30 (see Figure 4.4) with a hexahistidine tag attached at the N-terminus. The synthetic gene was introduced into *E. coli* BL21 or NM522, for recombinant expression. Reactivity of bacterially expressed Mk-TSP-2 against anti-tetrahistidine antibodies was detected by immunoblotting (Figure 4.6).

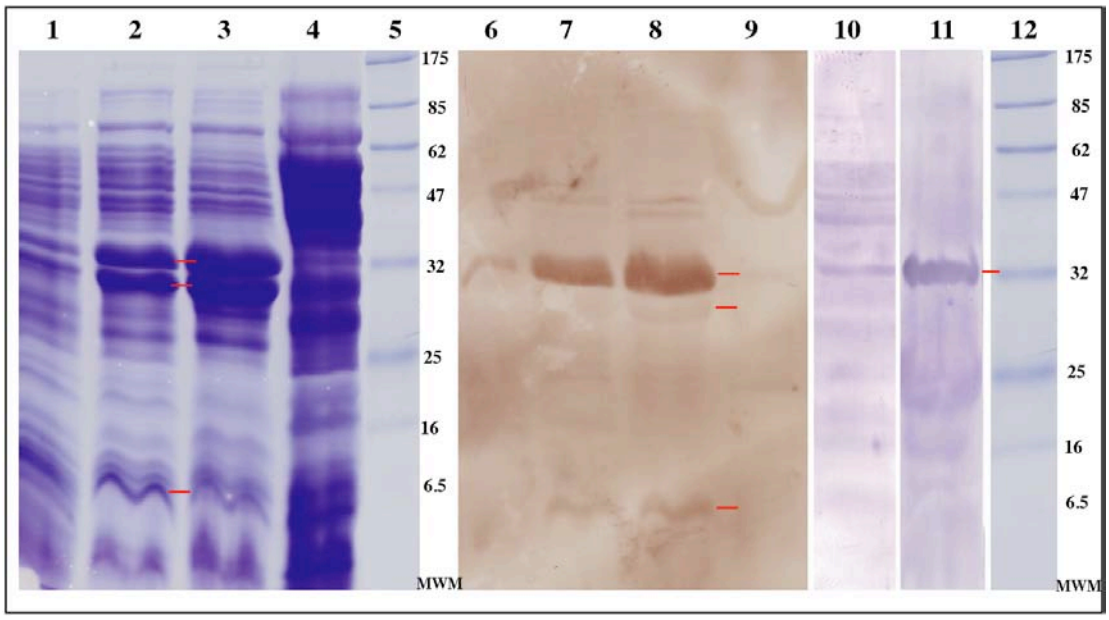


Figure 4.6 Expression of MK-ShK+TSP-2 and immunological affinity to species specific mice antibodies raised with Mk-venom proteins and CSL box jellyfish antivenom.

12.5% SDS-PAGE and western blot analyses of bacterially expressed Mk-TSP-2. Lanes 1 to 4 represent Coomassie blue-stained proteins corresponding to bacterial crude extract, bacterial pellet, sonication and supernatant, respectively. Lanes 6 to 9 show the reactivity of proteins corresponding to lanes 1–4 with anti-tetra his antibodies. The reactivity of the crude extract towards *M. kingi* specific mouse antiserum and CSL box jellyfish antivenom is shown in lanes 10 and 11, respectively. Molecular weight markers are shown in lanes 5 and 12.

In Figure 4.6, two bands of over-expressed proteins are clearly visible (Note the red indications) in lanes 1 to 3. Lane 4 corresponds to the supernatant and the lack of recombinant protein indicates that bacterially expressed Mk-ShK-TSP-2 is not soluble and remains as inclusion bodies in the pellet - lane 3 of Figure 4.6. Nevertheless, what is relevant is that the protein is being expressed and that the western blot analyses in lanes 6 to 8 (but not the soluble fraction in lane 9) demonstrate cross reactivity of the over-expressed protein with anti-tetrahistidine antibodies, although the doublet of over expressed proteins has differential immunological cross-reactivity. It is interesting to note that the upper band, with an apparent molecular weight of 32 Kda, is thicker and clearly displays stronger cross reactivity than the smaller band, suggesting that an N-terminal portion of the protein containing the histidine tag has been cleaved off. Furthermore, close examination of the primary structure revealed that there is the possibility of a pro-peptide cleavage in Mk-ShK+TSP-2 (indicated with a yellow line in figure 4.5 B). This was reinforced by the presence of a small protein band (< 6.5 kDa) displaying evident immunological reactivity against the anti-tetrahistidine antibodies. The molecular weight of the histidine tag together with the transit signal peptide and the predicted "propeptide" sequence theoretically is approximately 5.0 kDa, which correlates well with the migration pattern observed in the SDS-PAGE and western blot analyses. Since bacterial converting enzymes have relatively broad specificities and are considered analogous to their eukaryotic counterparts (Docherty & Steiner 1982), it is possible that MK-SHK+TSP-2 is releasing a fragment of the N-terminal SXC/ShK motif/domain *in vitro* by a bacteria processing protease. This would explain the doublet in which the upper band visibly displays more antigenicity, thus demonstrating the lack of the histidine tagged region in the lower band. At this point, it is crucial to mention that the N-terminus cleavage of ShK observed via SDS-PAGE and western blotting may occur as a result of the denaturing protein purification conditions. The presence of imidazole promotes protein unfolding and this could expose the predicted cleavage recognition site. In contrast, unfolding may not occur in native Mk-ShK+TSP-2 because the toxin domain has a compact and stable tertiary structure as a consequence of the six cysteine residues bonding three disulphide bridges and hiding the predicted propeptide cleavage site ending in KR. It has been reported that protein convertases can not recognize cleavage sites when they are present in compact structures stabilized by disulphide bridges, because the compact peptide folding prevent it (Kozlov & Grishin 2007).

Although immunoblot analysis of bacterial crude extract revealed the presence of a His-tagged protein band corresponding to the molecular weight of Mk-ShK+TSP-2, the recombinant protein did not exhibit a strong immunological reaction towards the *M. kingi*-specific mouse antibodies (Figure 4.6 Lane 10), thus potentially, indicating that TSPs may not be immunodominant antigens of Irukandji nematocysts. Nevertheless, it can not be ruled out that chymotrypsin-like serine protease is present in Irukandji venom and, in fact, its activity has been previously reported in other species of jellyfish venom (Gusmani *et al.* 1997). Another important consideration is that the ShK domain from TSP-2 was expressed independently of the peptidase domain and this did not exhibit antigenicity towards CSL box jellyfish antivenom. This was expected in advance due to primary structure analyses that indicated low antigenicity in that region. Of great significance is that TSP-2 cleaved from the ShK motif show strong cross-reactivity towards CSL antivenom, indicating that the antivenom contains TSP epitopes.

Mk-ShK+TSP-2 is the first protein from cubozoans to be successfully expressed in bacteria. Although optimization to obtain active recombinant Mk-ShK-TSP-2 from inclusion bodies needs to be carried out, the expression of Mk-ShK+TSP-2 is significant because this protease has an intriguing domain composition restricted to cnidarians. The N-terminal toxin homology domain is anticipated to be responsible for the expression in bacteria of Mk-ShK+TSP-2 by preventing the TSP domain from being active. It is possible that Cb-TSP-1 and Mk-TSP-1 did not express in bacteria at all because of the lack of this domain that serves as a protective peptide.

4.4.7 Gene expression pattern trypsin-like serine proteases in Cubozoa

Box jellyfish TSP genes were expressed differentially in adult medusa, as seen in Figure 4.7. The mRNA expression of Mk-TSP-1 (Figure 4.7 A) developed an early signal in clustered cells surrounding the nematocysts battery and 30 minutes later colour progressed in the gastric area (upon dissection, expression was also observed in the manubrium - data not shown). In contrast, Mk-TSP-2 (Figure 4.7 B) revealed an absence of gene expression in the gastric area but a clear signal restricted to clusters of nematocytes cells in the tentacles. No expression in

the gastric cavity, mouth or lip epithelia was observed after 24 hrs. The signal observed in these experiments is genuine due to the fact that not all body tissues were stained. For instance, no expression was observed in the exumbrella, pedalia region and tentacle areas in between nematocyst batteries of specimens treated with Mk-TSP-1 or Mk-ShK+TSP-2 riboprobes.

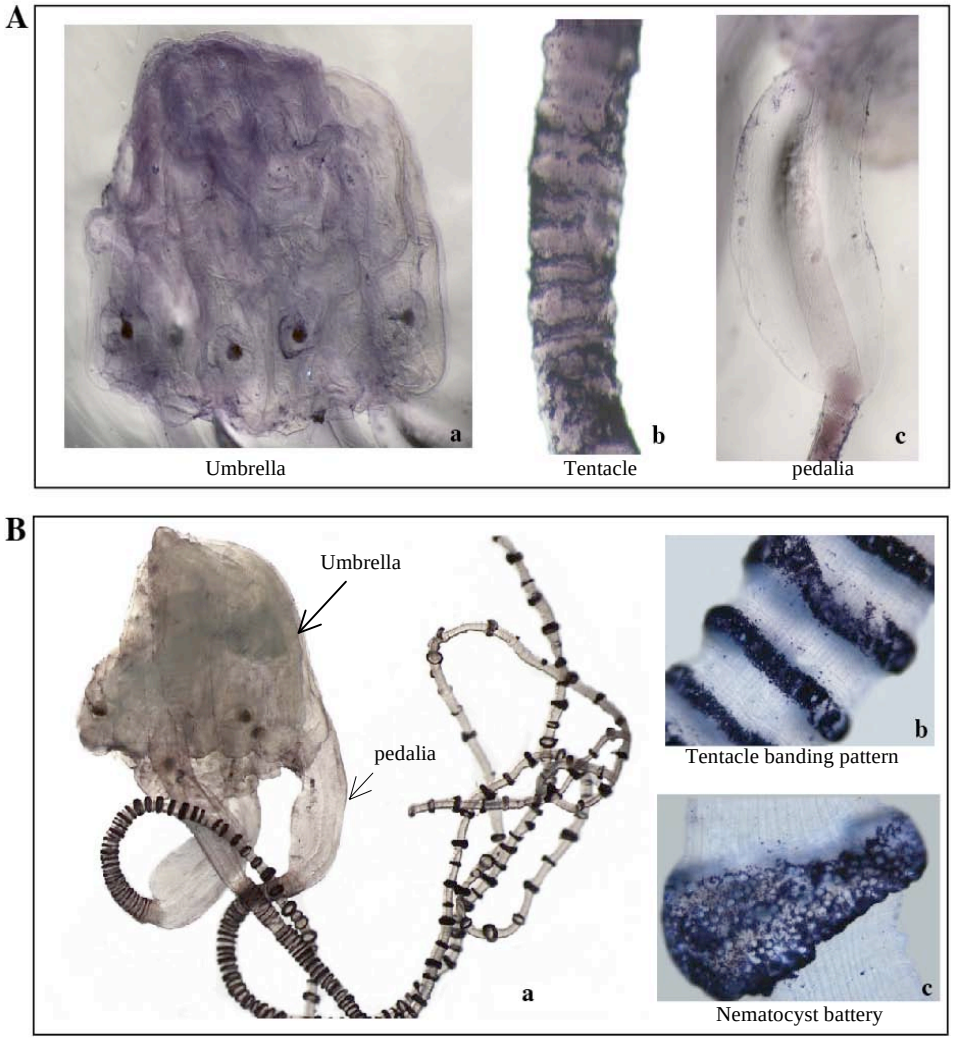


Figure 4. 7 *C. barnesi* whole mount *in situ* cross-hybridization of Mk-TSP-1 and Mk-ShK+TSP-2 mRNA probes.

Therefore, *in situ* hybridisation experiments suggest that both genes are associated with gastric digestion and at least Mk-ShK+TSP-2 is also likely to be involved with the venomous hunting apparatus of Irukandji jellyfish. Jointly, these molecular observations and the similarity of SXC/ShK motif with anemone neurotoxins, indicate that Mk-TSP-2 may play a dual role of

proteolytic digestion and neurotoxic activity. Similar suggestions have been previously reported in an astacin-like metalloproteinase containing a SXC/ShK toxin homology domain from the jellyfish *Podocoryne carnea* (Pan *et al.* 1998).

4.4.8 Comparative analysis of Cubozoa S1A enzymes

The relationship of the chymotrypsin family of serine proteases Group S1A (Barrett 2004) from Cubozoa and the known representative homologs retrieved from a BLASTP search were examined. The collapsed phylogenetic tree in Figure 4.8 shows three major clusters (A to C). Clusters B and C come from a common node and diverged. Clade C is the most diverse and includes vertebrate representatives and, at the basal branch, cnidarian and placozoan TSPs. Note that the vertebrate branches have longer distances indicating greater protein divergence in vertebrate TSPs. Similarly, cluster B is also dominated by vertebrates with a cnidarian representative as a sister branch. Cluster A contains exclusively cnidarians: anemones (*N. vectensis*), true jellyfish (*A. aurita*) and box jellyfish. Cluster A gives cnidarian TSPs an ancestral position which is in accordance with consensus phylogenetic views of cnidarian as the earliest divergent metazoa phyla.

Figure 4.8 B shows an expanded tree of the basal branch. Ascidians and sponges were used as an outgroup and two vertebrate representatives assisted in positioning the cnidarian cluster. Note this cluster is separated not only from the vertebrates but also from ascidian and sponge counterparts. Both approaches in Figure 4.8 A and B indicate no direct connection between TSPs in vertebrates versus invertebrates. It is clear that even between diploblastic animal (phylums such as Porifera and Cnidaria) or triplobastic animals (for example, chordates such as ascideans, fish and birds) do not have close phylogenetic relationships. For this reason it seems likely that TSPs in the animal kingdom have evolved independently having TSP-likprotein with different functions. The origin of S1A needs to be far more integrative because

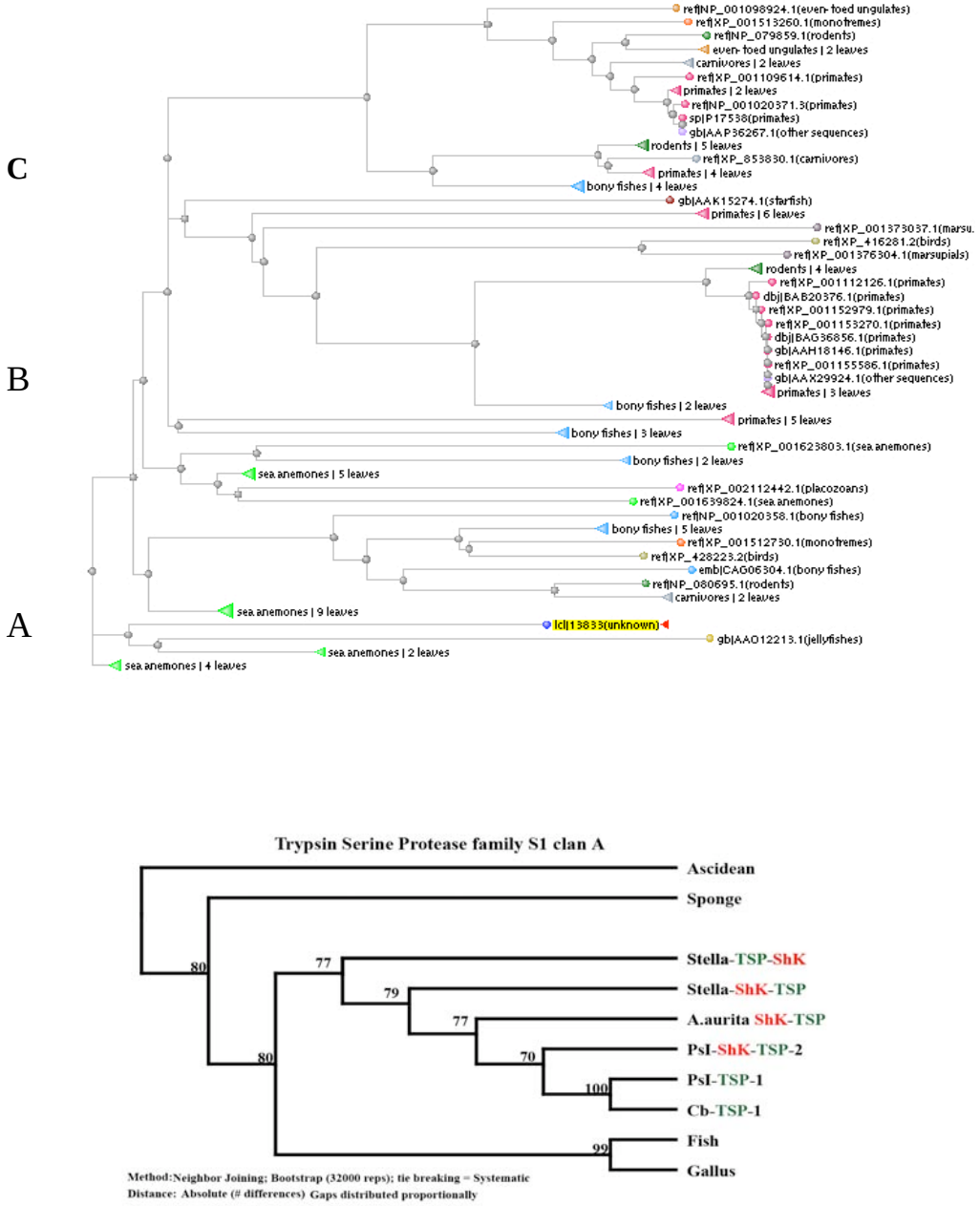


Figure 4.8 Phylogenetic relationship of trypsin-like serine peptidases and homologous species

A. NCBI distance tree (Fast minimum evolution, 0.85 of Maximum sequence difference) retrieved by NCBI BLASTP. B. The cluster basal is expanded including sponges and ascidians as an out group and two vertebrate representatives. Branch lengths represent the relative distance between sequence pairs determined by ClustalW.

this enzyme type is not just restricted to animals but there are also S1A representatives in fungi and bacteria. The last time the NCBI was interrogated in this study, new sequences from *Hydra magnipillata* and *Branchiostoma lanceolatum* (amphioxus=lancelet) were retrieved and these should be taken into consideration in future.

4.5 CONCLUSIONS

Thirty-three different cDNA clusters encoding peptidases were identified within the cubozoan EST collection. The peptidases found were members of four out of the five families identified so far in metazoans. The most highly expressed are serine proteases of the family S1A. Other interesting peptidases such as caspases and a threonine peptidase were also found. In addition, nine peptidase inhibitors were identified.

Two secreted chymotrypsin-like serine proteases were extensively studied: Cb-TSP-1/Mk-TSP-1 and Mk-ShK+TSP-2. The protein architecture and phylogenetic comparisons among three representative cnidarian proteases (NvTSP, Mk-ShK+TSP-2, *A. aurita* serine proteases) and other TSP-homologs in other metazoans species, revealed that although the serine protease domain is highly conserved across all animals, no trypsin serine protease-containing ShK domain is found in other metazoan fungi or virus. Therefore, it is concluded that the architecture observed in Mk-ShK+TSP-2 represents a cnidarian invention. Importantly, these non-mammalian serine proteinases have a high degree of similarity with human counterparts.

mRNA hybridisation studies indicated that Cb-TSP-1 and Mk-TSP-1 are involved in gastric digestion and also expressed in the nematocytes cells. In contrast, the Mk-ShK-TSP-2 ribo probe was absent in the gastro dermis and expression was restricted to clusters of nematocyte cells.

Elucidation of sequences for proteinaceous components from venomous jellyfish traditionally is restricted or unsuccessful because of the difficulties of obtaining sufficient biological material. In this sense, Mk-ShK+TSP-2, a defence protein, was expressed in *E. coli* as a histidine-tagged fusion protein in abundant quantities and displayed immunological affinity towards CSL box jellyfish antivenom. Although the recombinant product could not be obtained in a pure state, the aim of the study was directed towards successful expression, and

further experiments are required. The expression of Mk-ShK+TSP-2 recombinant protein may now allow full characterization of the protease activity of the TSP-2-ShK domain.

The combination of i) the comparative analysis of domain composition, ii) SDS-PAGE and western blot analyses of bacterially expressed Mk-ShK+TSP-2 showing strong cross-reaction towards *C. fleckeri* antivenom and iii) the surprisingly specific expression located exclusively in the nematocyte batteries of the tentacles may indicate that upon secretion, Mk-ShK+TSP-2 is stored in the nematocyst.

The ShK motif has an apparent propeptide cleavage site that may not be post-translationally affected because its primary structure indicates a compact motif. Therefore, perhaps there is no furin cleavage recognition unless the protein is denatured. Nevertheless, the NCBI protein domain DB identified in mature Mk-ShK+TSP-2 a cleavage site that is likely to be cut off in order to activate the TSP domain. It is unknown at this stage whether the propeptide containing the ShK domain homologous to anemones neurotoxic peptides may have a role in paralysing the prey.

This is the first report of genes related to proteolysis or proteolysis inhibition that have been reported for cubozoans.

CHAPTER 5

The ShK-like toxin family in box jellyfish. Molecular identification, protein and gene expression studies

5.1 ABSTRACT

The occurrence of *CbTXI and Mk-332,* two orthologous gene members of the ShK (*Stichodactyla helianthus* K^+ *toxin)* family from *C. barnesi and M. kingi* is reported. Since each gene has two isoforms, four cDNAs were characterized. The proteins are synthesized as pre-propeptide precursors. Protein maturation involves post-translational cleavage at a site ending with a pair of basic residues, lysine and arginine (KR). The precursor organization in these cubozoan genes is similar to that for short chain K^+ neurotoxins, Na^+ neurotoxins (Moran *et al.* 2008a), pore forming toxins from the nematocyst of anemones, nematocyst collagens (Anderluh *et al.* 2000b) and "Aurelin", an antimicrobial protein from the scyphomedusan, *Aurelia aurita* (Ovchinnikova *et al.* 2006).

The prepro regions either do not contain cysteine, or may have one cysteine residue. The mature proteins have a tandem arrangement of three to four cysteine rich motifs which are homologs of short chain K^+ channel toxins of sea anemones. Invariably, in each domain the first four residues have a distinctive X**C**X**D** pattern and three out of the six cysteines are found in the last eight residues with the signature **C**XXX**TC**XX**C** (Pan *et al.* 1998)

Two of the proteins were bacterially expressed. Preliminary functional assays using cockroaches clearly indicated neurotoxic activity. Protein domain architecture, gene expression studies and functional assays indicated the involvement of these molecules in the defence system of box jellyfish. This is the first report on defence preproproteins from Irukandji box jellyfish.

Key words*:* Alternative splicing, alternative polyadenylation signals, secreted propeptides rich in cysteine.

5.2 BACKGROUND

Box jellyfish, like other cnidarians such as sea anemones, corals and hydra, are soft bodied animals with either sessile or free leaving forms, which depend on sophisticated offensive/defensive mechanical gadgets (nematocytes) combined with a complex molecular armament. This armament has developed based on selection pressure to survive a predatory environment, where predator deterrence and prey capture traits are important factors in competitiveness.

Cnidarian bioactive components have been identify throughout the entire organism in tissues rich in nematocytes, mesoglea (Ovchinnikova *et al.* 2006), tentacles and whole body secretions (Aneiros & Garateix 2004) *e.g* ectodermal mucus. The chemical arsenal includes lethal cytolysins, neurotoxins (e.g. classic short chain neurotoxins affecting sodium and potassium conductance), phospholipases (Talvinen & Nevalainen 2002), proteases (Calton & Burnett 1982; Gusmani *et al.* 1997; Tamkun & Hessinger 1981), protease inhibitors (Antuch *et al.* 1993; Delfin *et al.* 1994) and antimicrobials (Morales-Landa *et al.* 2007; Ovchinnikova *et al.* 2006). Despite the fact that several cnidarian toxic proteins or peptides have already been thoroughly characterized with regards to gene sequence, protein architecture and function, only five lethal cytotoxins of high molecular weight have been identified, cloned and characterized so far from cubozoans (Brinkman & Burnell 2007; Nagai 2003).

The research covered in this chapter was designed to determine whether a homolog of ShK toxin from anemones (Castaneda *et al.* 1995; Cotton *et al.* 1997) was present in cubozoans, and if so, to further characterize and compare the gene. The purpose of identifying and characterizing these genes within Irukandji cDNA libraries was due not only to the putative neurotoxic properties of the gene products, but also because the ShK domain may have been recruited and diversified into multi domain proteins with different functions within cnidarians as well as in other more evolved metazoans. The relevance of having molecular data on genes involved in defence mechanisms from the early diverging Cnidaria phylum like may allow us to uncover ancestral roles and understand the evolutionary history of metazoan gene families involved in predation and foe deterrence.

5.3 EXPERIMENTAL PROCEDURES

RNA isolation and library construction was outlined in Chapter 2 and the expressed sequence tag collection is described in Chapter 3.

5.3.1 Gene identification and isolation of full length clones by cDNA hybridization.

pBluescript SK- (pBSK-) universal primers were used in combination with forward, reverse, and reverse complementary primers designed based on the most frequent codon usage in cubozoans to target the amino acid sequence QLGLVSWGSGCAQK:

```
5′CAATTGGGATTGGTTTCTTGGGGATCTGGATGTGCACAAAAA 3′
   Q  L  G  L  V  S  W  G  S  G  C  A  Q  K
```

Purified λ DNA (1:1000) from the *C. barnesi* cDNA library was used as template in a gradient PCR using a combination of the reverse complementary primer (sequence underlined above) and the forward pBSK- universal primer, T3. Reactions were performed using GoTaq® Flexi DNA Polymerase (Promega) according to the manufacturer's instructions and under the following parameters: 95ºC for 15 minutes, 35 cycles of 95ºC for 1 min, 47.5 ºC for 1 min (+0.5 ºC each cycle), 72ºC for 1 min and 72ºC for 5 min. The reaction product was resolved by agarose gel electrophoresis (1% w/v), the gel portion ranging from 300 to 800 bp excised, the DNA gel-purified and used as template for another round of PCR. PCR was performed in 50 µL containing 1.25 units of GoTaq® Flexi DNA Polymerase, 3.0 mM Mg^{2+}, 10 ng DNA, the same set of primers (0.1µM) and the Tris-based buffer containing KCl supplied by the manufacturer (Promega). The cycling conditions included a denaturation step at 95ºC for 5 min, 30 cycles of denaturation at 95ºC for 45 sec, annealing at 55ºC for 45 sec, extension at 72ºC for 1 min and a final extension at 72ºC for 5 min. A single 356 bp band from this secondary PCR was obtained and the purified PCR product cloned into pGEM-T™ and sequenced. The CbTXI fragment was identified by nucleotide sequencing.

CbTXI full length clones were obtained by screening the cDNA library. A DNA probe was generated by PCR using a forward [(G42) 5′-GCTTCTGGAAAGGACAAGG-3′] and a

reverse primer [(G43) 5′-GTTCACTCGTTGCCTGTTG-3′], and cycling conditions of 95ºC for 5 min, 32 cycles at 95ºC for 30 sec, 55ºC for 30 sec, 72ºC for 1 min and 72ºC for 1.5 min. 25 ng of purified PCR product was radiolabelled with [α-^{32}P]dATP using random primers (DECAprime II kit, Ambion). Plaques (6 x 10^6) of the *C. barnesi* cDNA library were screened, positive plaques identified and inserts confirmed by PCR, using gene-specific primers and pBSK- universal primers. Plaques were re-screened and single, positive plaques isolated. Plasmid DNA was excised using the *in vivo* excision strategy (Stratagene) and plasmid inserts sequenced in both directions (Macrogen, Korea).

MK-332 isoform I was a truncated 5′-end tag identified by analysis of the *M. kingi* EST collection. A full length sequence was obtained by DNA hybridization using a labelled [α-^{32}P] PCR-generated probe using the forward primer (G45) 5'-GGATGCTGCTGGGATTTTCAGACTC-3' and the reverse primer (G46) 5'-CCTTCTCCACATACTCCACAAGTTCTCG-3'. The cycling conditions used were as follows: denaturation at 95ºC for 5 min, 32 cycles of 95ºC for 1 min, 55ºC for 1 min, 72ºC for 1.5 min, and a final extension at 72ºC for 5 min.

Protein motifs were identified by searching at http://smart.embl-heidelberg.de/smart, BLASTP. Genome searches were undertaken at http://genome.jgi-psf.org/.

5.3.2 Cloning of expression constructs and fusion protein expression

Expression constructs were designed to contain 5′-*Sph*I and 3′-*Kpn*I or 5′-*Sph*I and 3′-*Hind*III sites to facilitate in-frame cloning into pQE30 (Qiagen). Primers used are listed in Table 5.1.
The expression plasmids were constructed using a two-step PCR strategy (cycling times and temperatures were chosen according to standard parameters based on the nucleotide composition of the primers). The first PCR product was cloned into pGEM-T^{TM} (Promega) and sequenced. Following confirmation of clone identity by DNA sequencing, plasmids were double-digested with *Sph*I/*Kpn*I or *Sph*I/*Hind*III (New England Biolabs) and ligated into pQE30. Successful ligation of the insert into pQE30 was confirmed by DNA sequencing.

Table 5. 1 Primers used for expression constructs

The left column indicates gene name, followed by the forward and reverse primer sequences, including restriction sites.

Gene name	Forward primer	Reverse primer
CbTXI	G58 Sph I 5'-GCATGCATTTCTATACCCTTACGGCTG-3'	G60 HindIII 5'-AAGCTTACACAAGTTGCATGTTTTCTTAC-3'.
CbTXII	G59 SphI 5'-GCATGCCGGAGGAAAATACACCGATG-3'	G60 HindIII 5' AAGCTTACACAAGTTGCATGTTTTCTTAC-3'.
MK-332-II	G61 Sph I 5' GCATGCGCTGTTGCCAAAGTATTAG-3' G611 Sph I 5′-GCATGCGCAATAGATGCTTCATG-3′	G62 HindIII 5′-AAGCTTGCAAATACCGCAAGTTTTCCTG-3′

For recombinant expression, 5 mL overnight cultures of transformed bacteria (*E. coli* NM522 or BL21 strains) were used to inoculate 500 mL LB medium containing 50 $mg.mL^{-1}$ carbenicillin in a baffled 2 L conical flask, which was shaken at 37ºC until the culture reached the mid-exponential growth phase (OD_{600}=0.4-0.6). Fusion protein expression was induced by the addition of IPTG to a final concentration of 1 mM and cultures shaken at 30ºC for 12 hrs before bacteria were harvested by centrifugation. Recombinant protein was analyzed by SDS-PAGE (12.5 % w/v polyacrylamide) and western blotting. The expressed protein in the soluble fraction was purified from the bacterial lysate in two steps (batch purification followed by column chromatography) using non-denaturing conditions and a metal affinity resin (Qiagen)**.**

5.3.3 Preliminary functional assay

Purified recombinant protein was dialyzed against PBS (6 hrs and overnight following replacement of PBS). Approximately 0.1 µg in 0.5 µL of PBS was injected into the third integument of cockroaches; negative controls contained PBS only. This experiment is not statistically represented.

5.3.4 in situ hybridization

mRNA *in situ* hybridization was conducted on jellyfish specimens as described in Chapter 4 using digoxigenin-labelled riboprobes created with the full length sequences of CbTX-I and Mk-332 isoform I. Juvenile and adult specimens were used.

5.4 RESULTS AND DISCUSSION

5.4.1 CbTX-I and CbTX-II gene identification, isolation and characterization

Crude extracts of "*C. barnesi*" were analysed by SDS-PAGE and western blot analysis using CSL box jellyfish antivenom (Wiltshire *et al.* 2000) and a protein band was identified to be a putative neurotoxin (Winkel; personal communication from unpublished data). A N-terminal peptide sequence (QLGLVSWGSCAQK) was obtained and used in this study to design a set of oligonucleotides to amplify clones from the *C. barnesi* cDNA library. An amplicon of 379 bp, named *pcr*CbTXI, was generated. The primed region has 67 % similarity at the nucleotide level and 100% identities in the first ten nucleotides of the primer used in correlation with the sequence obtained. The nucleotide sequence primed does not exactly encode the residues expected but has 50% similarity with the peptide predetermined as the target, in this case a putative neurotoxin.

A

```
Identities = 7 (50%)
Provided amino acid sequence      1 QLGLVSWGSGCA-QK  14
pcrCbTX encoded product         270 TCGLGNGG-GCADQP 283
                                     GL   G GCA Q
```

B

```
Identities = 19 (67%)
CbTX Rev/Comp primer              1 GGATTGGTTTCTTGGGGATCTGGATGTG 28
CbXI                              1 GGATTGTGTAATGGTGG---TGGATGTG 25
                                    ******  *  * * **   ********
```

Figure 5. 1 Amino acid and nucleotide targeted and experimentally resulted sequences

In A, the top sequence is the N-terminal sequence obtained by Wiltshire *et al.* (2000) and the lower sequence is the primed encoding region. In B, the top sequence is the oligonucleotide designed according to the amino acid sequence QLGLVSWGSCAQK and the preferred codon usage in cubozoans.

The dissimilarities in protein sequences between the target and the primed encoding sequence may be due to artifacts generated during amino acid sequencing, however, the more likely explanation is the possibility of protein variation between species. In other words, it is possible that the target does not come from *C. barnesi* but other Irukandji species in the North QLD region, particularly in Palm Cove. This scenario is plausible, because the nematocyst and overall features described in samples collected at Palm Cove and used by Wiltshire and

colleagues (2000), do not correlate with the description of *C. barnesi* nematocysts (Gershwin 2006b). It was only recently observed that more than one *Carukia sp.* (*Carukia horeae* and *C. barnesi*) inhabit the same geographic area. Nonetheless, the nucleotide and translated sequence used for PCR primers is given in Figure 5.1 A. The 379 bp PCR product consisted of a 138 bp ORF and 241 bp of 3′-UTR ending with a poly A tail. Significantly, the translated sequence contained in the last eight residues the signature **C**XXX**C**XX**C**, classically found in peptidic neurotoxins acting on voltage-gated K^+ channels.

The amino acid sequence encoded by *pcr*CbTXI was analyzed and the question was raised as to whether *pcr*CbTXI represented an incomplete but homologous gene to the short chain neurotoxins previously identified in anemones. Examples include the *S. helianthus* K+ toxin, ShK (Castaneda *et al.* 1995), or HmK from *Heteractis magnifica* (Gendeh *et al.* 1997a). Thus, an attempt was made to obtain the full-length clone from the *C. barnesi* cDNA library by standard DNA hybridization methods using a radiolabelled DNA fragment of 134 bp. Based on homology with previously characterized ShK-like toxins, a number of clones within the range of 450 bp to 1,350 bp were selected and sequenced. Sequencing results indicated that clones smaller than 984 bp were truncated cDNAs of two transcript populations named CbTX-I and CbTX-II. The nucleotide and deduced amino acid sequences of CbTX-I and CbTX-II are provided in Figure 5.2 and 5.3 respectively.

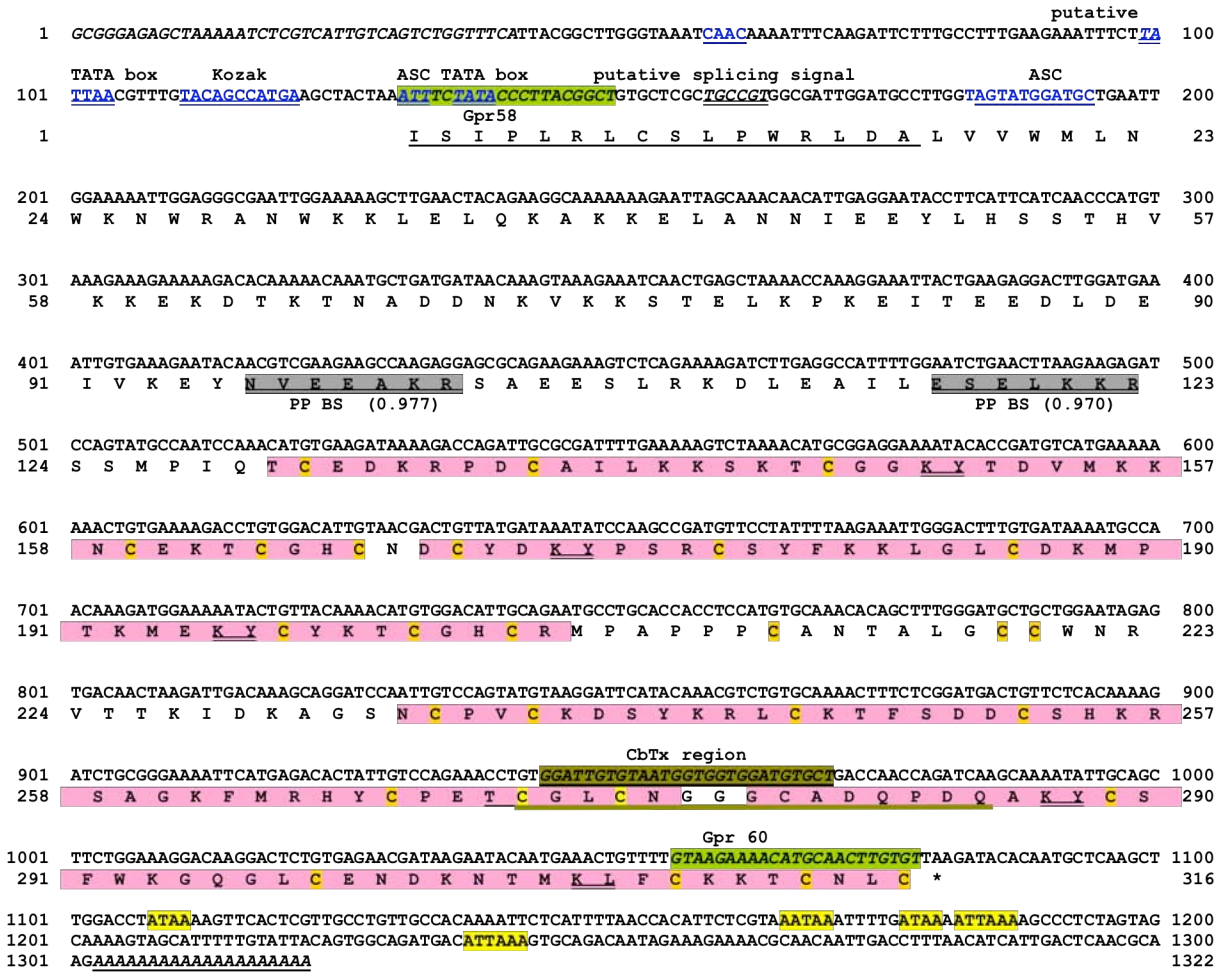


Figure 5. 2 Nucleotide and deduced amino acid sequence of the CbTX-I proprotein precursor

Nucleotide and deduced amino acid sequence of CbTX-I cDNA preproprotein precursor. Promoter elements are indicated in blue. The sequence used for reverse primer binding is shaded in dark green. The primer binding sites used for engineering the protein expression plasmid construct are shaded in bright green. The secretion peptide is underlined in black and is followed by a prosequence containing two putative proconvertase binding sites (PP BS) shaded in grey. Prop 1.0 Server scores are shown in brackets. The region encoding the mature protein includes four ShK toxic domains, which are shaded in pink. Polyadenylation signals (5) are higlighted in yellow.

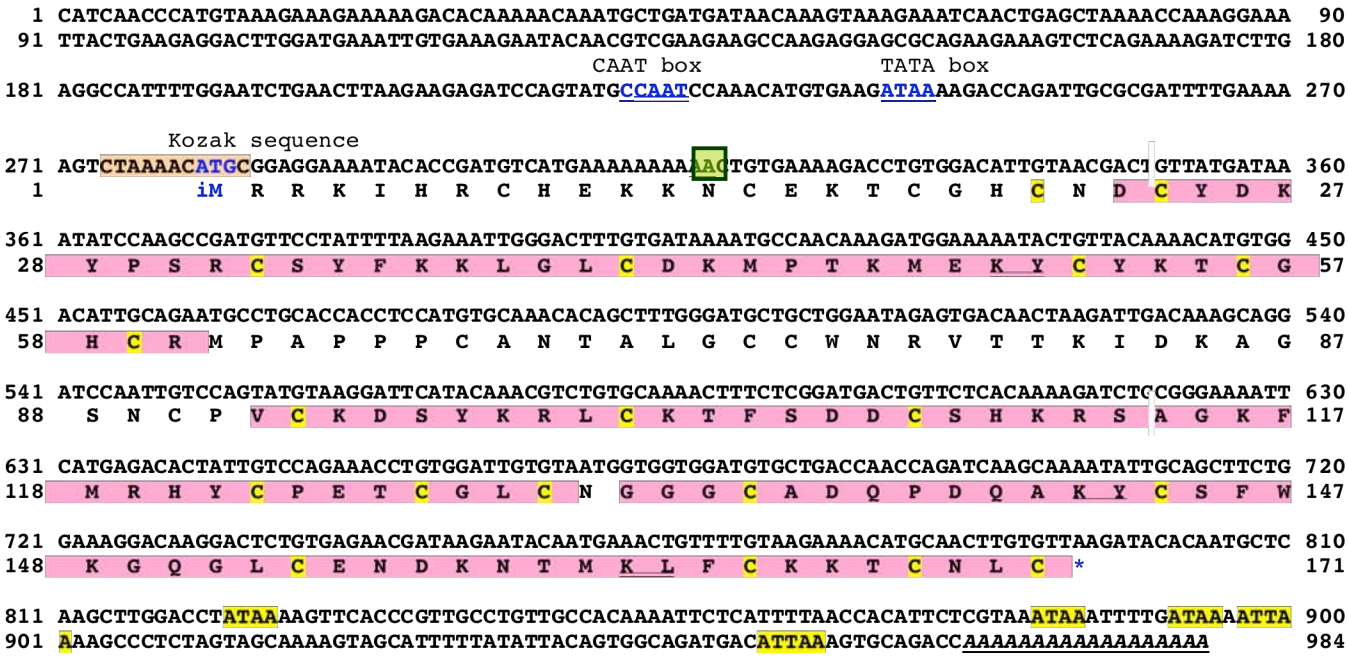


Figure 5. 3 Nucleotide and deduced amino acid sequence of the CbTX-II proprotein precursor.

Promoters elements, initiation of translation site are indicated in blue. Two adenines highlighted in green indicate an insertion that is not observed in CbTX-I. Cysteine residues are highlighted (yellow) and the toxin domains highlighted in pink. At the 3′-UTR, five polyadenylation signals are indicated in yellow.

The full-length CbTX-I transcript is 1,322 bp with the first 131 bp corresponding to the 5'-UTR, followed by 948 bp of coding region and 242 bp of 3'-UTR with a poly A tail (Figure 5.2). The 5′-UTR contains a promoter region with a putative TATA box and a translation initiation codon within a sequence having 72% similarity to the Kozak consensus sequence: GCCAGCCAUGG (Kozak 1989). However, the AUG start codon is not in frame and 10 nucleotides upstream there is an AUU codon (isoleucine), which could be an alternate start codon. Preceding the AUU codon, there is also an alternative promoter element (TATA box) and a third transcription initiation site. Although rare, translation in eukaryotes can be initiated

by codons other than AUG. These include the codons CUG, GUG, UUG, AUA, ACG and AUU (Wegrzyn *et al.* 2008). Significantly, alternate transcription initiator regions with characteristic transcriptional regulatory elements including CAAT and TATA box consensus sequences have been reported for cnidarian toxins that modify K^+ channels (Gendeh *et al.* 1997a). Interestingly, within the 3′-UTR there are five polyadenylation signal sites upstream of the poly A tail. The use of an alternate transcription start site and multiple polyadenylation signal sites could be important in the regulation of toxin-producing cells according to stress levels.

A predicted splice signal in the region that encodes the signal peptide of CbTX-I is marked in Figure 5.2, and in comparison, an almost identical splice signal is found in Nv1, a short neurotoxin that inhibits the inactivation of voltage gated sodium channels from *N. vectensis* (Moran *et al.* 2008a).

The ORF translated into a preproprotein precursor of 316 amino acids. Hydrophilicity plot analysis of CbTX-I identified a 17-residue hydrophobic region situated at the N-terminus, which correlated with the SignalP 3.0 server prediction for the most likely cleavage site between residues 16 and 17 (LDA-LV). Preceding the pre-sequence, there is a pro-sequence of 107 amino acids (12.6 kDa), containing two protease cleavage recognition sites ending with lysine and arginine in positions 101-102 and 122-123. Both predictions retrieved high scores and both are located before the signal peptide, the first ShK motif and the first cysteine residue. Moreover, the motifs for processing satisfy the basic cleavage control rule and contain arginine (Kozlov & Grishin 2007). However, the existence of two proconvertase binding sites so close together is unlikely. Nonetheless, the presence of the signal peptide and the pro-sequence indicates that CbTX-I is a secreted protein that requires post-translational processing (Figure 5.2). The mature protein consists of 194 amino acids, with a calculated molecular weight of 21.67 kDa. It is rich in cysteines (28) and lysines (28). Analysis of the amino acid composition indicates the protein is polar, with an estimated pI of 9.75.

In contrast, CbTX-II represented in Figure 5.3, is a 984 bp cDNA clone which is composed of 278 bp 5'-UTR, 516 bp coding region, 189 bp 3'-UTR and five polyadenylation signals upstream of a poly A tail. The 5′-UTR contains promoter elements such as a CAAT box and a putative TATA box alongside a Kozak sequence CTAAAACATGC with 45.5% similarity

with the consensus sequence GCCAGCCAUGG (Kozak 1989). The cDNA ORF translated into a polypeptide of 172 amino acid residues.

At the protein level, CbTX-II has the same primary structure as CbTX-I, however the N-terminal sequences differ. The nucleotide sequence of CbTX-II contains a two base insertion (both adenines) at nucleotides 309 and 310, which leads to a frame shift and translates into a polypeptide that does not contain a classic transit signal peptide or a proconvertase cleavage region. No other insertions or deletions were observed within the ORF. Four reads of each base were obtained from different clones in order to rule out the possibility of a sequencing error.

CbTX-II is a polar protein with an isoelectric point of 9.79 and an estimated molecular weight of 18.16 kDa. The protein contains 25 cysteines and 23 lysines making up 29.82 % of the protein. In comparison to CbTX-I, three not four ShK domains with the classic six cysteine signature are contained in CbTX-II.

According to NTXProtein, a neurotoxic protein prediction database hosted at http://www.imtech.res.in/raghava/ntxpred (Saha & Raghava 2007), CbTX-II could be active on postsynaptic membranes, playing a role in producing peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. It is predicted to bind to the nicotinic acetylcholine receptors, although the putative prediction of this protein needs to be further analysed by functional assays using the recombinant protein.

Comparisons made of the 3'-UTRs (Figure 5.4) support the fact that CbTX-I and CbTX-II come from two different transcript populations. Specifically, in CbTX-I there are 52 nucleotides inserted upstream of the poly A tail. Both transcript isoforms harbour five polyadenylation signal sequences upstream of the respective poly A tails, indicating alternative usage of the polyadenylation signal. Similarly, multiple polyadenylation signals have been detected in the antimicrobial protein Aurelin from the scyphomedusa, *A. aurita*. Aurelin, which has defensin-like features, also contains an ShK-like toxin homology domain (Ovchinnikova *et al.* 2006). Moreover, variable 3′-UTRs and differential usage of polyadenylation signals have also been found in scorpion venom neurotoxins acting on K^+ channels (Zhu & Tytgat 2004).

```
CbXI     1 GATACACAATGCTCAAGCTTGGACCTATAAAAAGTTCACTCGTTGCCTGTT  50
CbXII    1 GATACACAATGCTCAAGCTTGGACCTATAAAAAGTTCACCCGTTGCCTGTT  50
                                                 *

CbXI    51 GCCACAAAATTCTCATTTTAACCACATTCTCGTAAATAAATTTTGATAAA 100
CbXII   51 GCCACAAAATTCTCATTTTAACCACATTCTCGTAAATAAATTTTGATAAA 100

CbXI   101 ATTAAAAGCCCTCTAGTAGCAAAAGTAGCATTTTTGTATTACAGTGGCAG 150
CbXII  101 ATTAAAAGCCCTCTAGTAGCAAAAGTAGCATTTTTATATTACAGTGGCAG 150
                                                *
CbXI   151 ATGACATTAAAGTGCAGACAATAGAAAGAAAACGCAACAATTGACCTTTA 200
CbXII  151 ATGACATTAAAGTGCAGACCAAAAAAAAAAAAAAAAAA----------- 188
                              * * *   *    ***

CbXI   201 ACATCATTGACTCAACGCAAGAAAAAAAAAAAAAAAAAAA 241
CbXII  189 ---------------------------------------- 188
```

Figure 5. 4 Alignment of the 3′-UTR of CbTX-I and CbTX-II.

Nucleotide dissimilarities are indicated by asterisks in the consensus line. Polyadenylation signals are shaded in yellow.

It is possible that CbTX-I and CbTX-II are generated from the same gene by alternative splicing and RNA editing, resulting in two transcript variants encoding different proteins. However, it is also possible that they are encoded by different genes within a multicopy family. Further analyses by Southern and northern blotting are required to clarify whether these mRNAs originate from a single gene or result from gene duplication events. It is anticipated that northern blotting analyses would yield multiple bands with a risk of misinterpretation due to the presence of five polyadenylation signals upstream of the poly A tail.

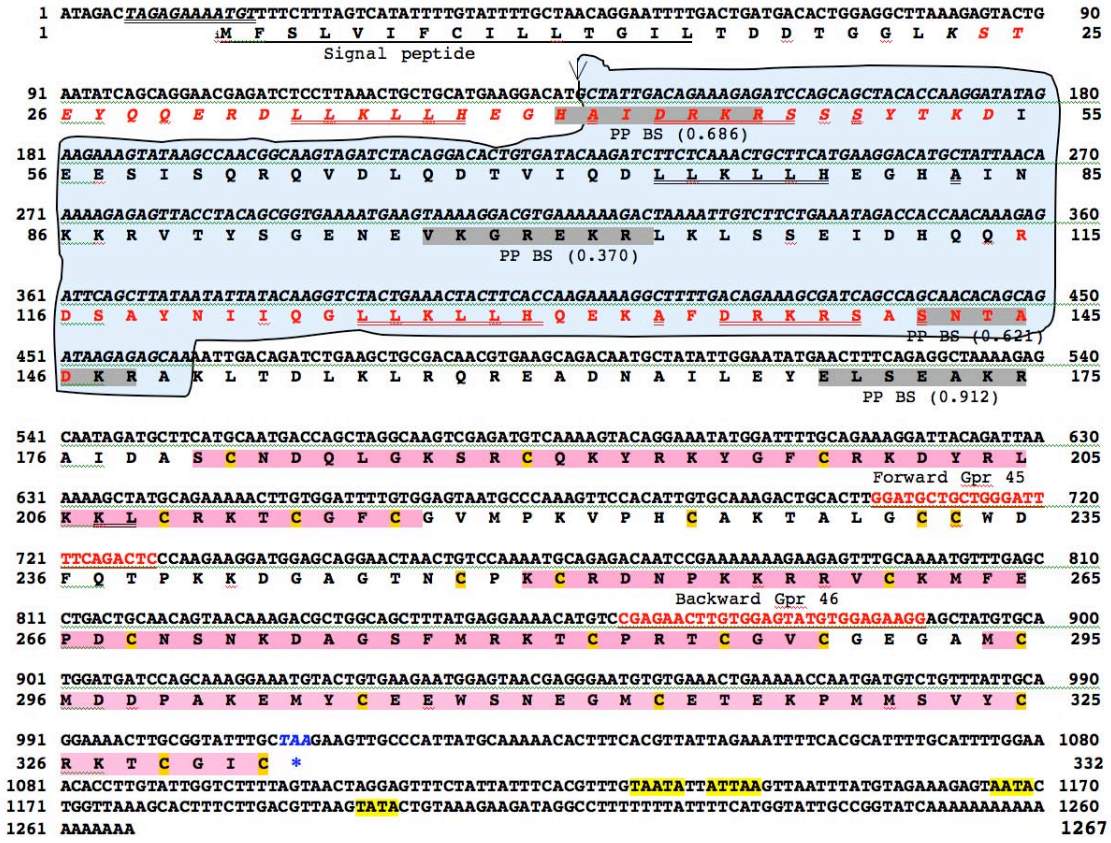
5.4.2 Mk-332-I and II gene identification, isolation and characterization

The original clone corresponding to Mk-EST-332 was truncated. Full length cDNAs were obtained by hybridization of the *M. kingi* cDNA library using a 185 bp radiolabelled probe. Sequencing of positive clones allowed the identification of two splice variants or isoforms, named Mk-332-I and II (Figures 5.5 and 5.6).

Mk-332-I is 1,267 bp, of which 1,008 bp correspond to coding region followed by 237 bp of 3′-UTR that does not contain any distinctive domains and which ends with a poly A tail. The coding region encodes a precursor of 336 amino acids with a signal secretion peptide of 15 amino acid residues. The pro-sequence located between the signal peptide and the mature protein is 160 residues in length. Two internal repeats of 93 bp with 63% identity were

identified via DNA matrix alignments. The amino acid products have 51% identity and 19% similarity. Repeat sequences are indicated in red and nucleotide and amino acid alignments are also provided in Figure 5.5.

A



B

Figure 5. 5 Schematic representation of the Mk-332-I cDNA and translated precursor

A. The blue area marks the sequence region (325 bp) missing in isoform II. Underlined amino acid residues indicate a repeating pattern. The initiating methionine is included in the underlined secretion signal peptide and the stop codon is highlighted in blue. Four possible proprotein processing binding sites (PP BS), with scores retrieved by Prop 1.0 Server in brackets, are shaded in grey. Three ShK domains are shaded in pink. Cysteine residues are highlighted in yellow. The 3′-UTR has several alternative polyadenylation signals, highlighted in yellow.

B. Nucleotide sequence alignment corresponding to the repeat regions within the propeptide region

Mk-332 isoform 2 (Mk-332-II) is represented in Figure 5.6. Mk-332-II is 957 bp including 12 bp of 5′-UTR and an AUG codon present within a Kozak sequence that has 54.5% similarity to the consensus sequence (Kozak 1989). There are 648 bp of coding region, from an initiating methionine to the terminating codon, TAA. The 3′-UTR is 271 bp and includes seven nucleotides before the poly A tail that are absent in Mk-332-I. The ORF translated into a precursor of 216 amino acid residues. The first 27 residues correspond to a signal peptide with a cleavage site between positions 25 and 26. There is a preproconvertase binding site (PPC BS) that produces a propeptide of 33 residues (3.4 kDa; pI = 4.08.). The mature protein contains 158 amino acid residues with an approximate molecular weight of 17.9 kDa and an estimated pI of 9.89.

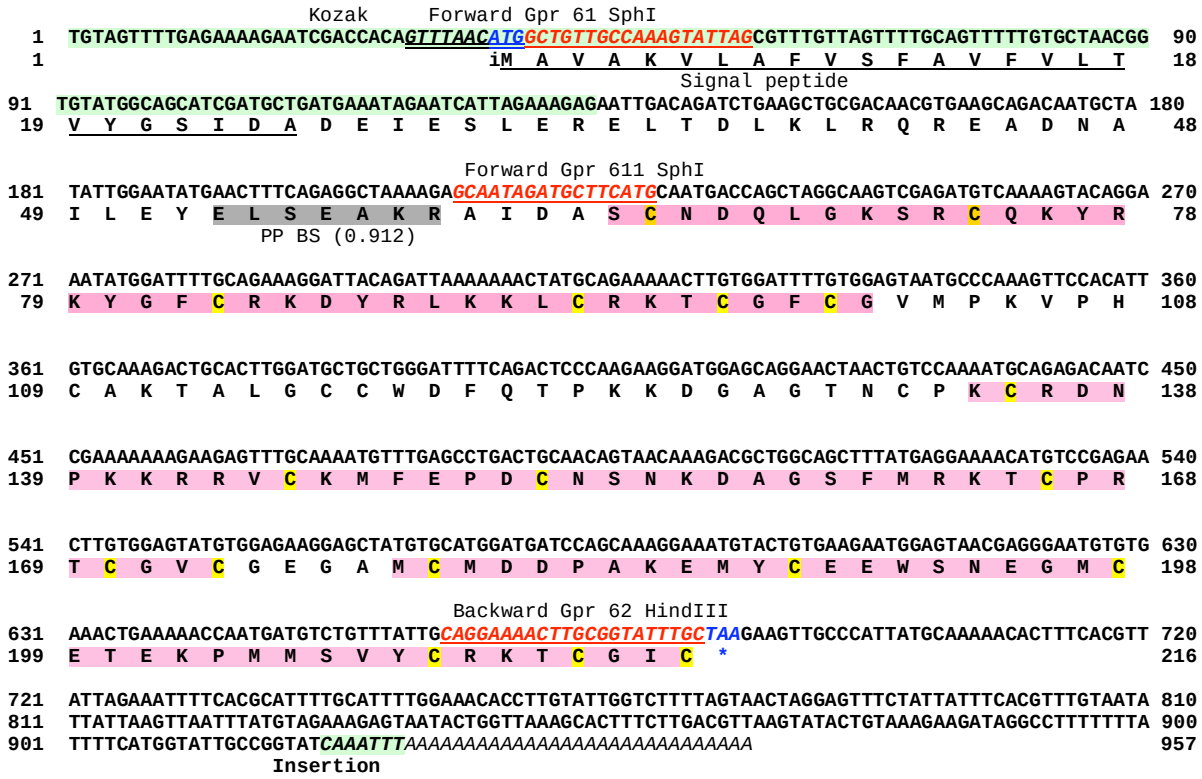


Figure 5. 6 Schematic representation of the Mk-332 splice isoform 2 cDNA (Mk-332-II) and translated precursor

Primers used for cloning into the expression vector are indicated in red without the restriction enzyme site sequences. The region in green at the 5′-end has 47% similarity with the first 137 residues of Mk-332-I. A small insertion before the poly A tail that is not observed in isoform I is also indicated in green. The signal peptide is underlined. A proconvertase recognition site (PP BS), with the score retrieved by Prop 1.0 Server in brackets, is indicated in grey. Three ShK domains are highlighted in pink. Cysteine residues are highlighted in yellow.

The prosequence of Mk-332-II has four predicted proprotein processing sites ending with a lysine-arginine doublet similar to CbTX-I. The predicted mature protein contains 157 amino acid residues. According to the amino acid composition the calculated molecular weight is 17.75 kDa with an estimated isoelectric point of 9.82. The most abundant residues in the protein are cysteine (22), lysine (20), glycine (13) and arginine (11), representing 14%, 12.7%, 8.2% and 11% of the total protein, respectively. The polypeptide encodes three ShK toxin domains.

An alignment of Mk-332-I and Mk-332-II is shown in Figure 5.7. The 5′ region of Mk-332-I contains two sequences of 138 nucleotides, which have 47% similarity. These repeats are found within the first 436 nucleotides of Mk-332-I. Interestingly, the first 138 bp of isoform II has 47 and 37% similarity with the repeats in Mk-332-I, respectively. In Figure 5.7, these regions are underlined. The first repeat is between nucleotides 1 to 138, and the second repeat is between nucleotides 328 to 462. The DNA similarities do not translate into protein similarities in that region.

```
332-1     1 ATAGACTAGAGAAAATGTTTTCTTTAGTCATATTTTGTATTTTGCTAACAGGAATTTTGACTGATGACACTGGAGGCTTAAAGAGTACTG    90
332-2     1 ----------------------------------------------------------------------------------------     0
             ***  * *******   *           **   **  * **** ** ** * *   *  ** *     ** **  **   *   ** *

332-1    91 AATATCAGCAGGAACGAGATCTCCTTAAACTGCTGCATGAAGGACATGCTATTGACAGAAAGAGATCCAGCAGCTACACCAAGGATATAG   180
332-2     1 ----------------------------------------------------------------------------------------     0
               ***  **** * ***   **  * ***  *   *** *   * *

332-1   181 AAGAAAGTATAAGCCAACGGCAAGTAGATCTACAGGACACTGTGATACAAGATCTTCTCAAACTGCTTCATGAAGGACATGCTATTAACA   270
332-2     1 ----------------------------------------------------------------------------------------     0

332-1   271 AAAAGAGAGTTACCTACAGCGGTGAAAATGAAGTAAAAGGACGTGAAAAAAGACTAAAATTGTCTTCTGAAATAGACCACCAACAAAGAG   360
332-2     1 ----------------------------------------------------------TGTAGTTTTGAGAAAAGAATCGACCACAGTTTA    33
                                                                       * ** **  ****   * *  * ***

332-1   361 ATTCAGCTTATAATATTATACAAG-GTCTACTGAAACTAC--TTCACCAAGAAAAGGCTTTTGACAGAAAGCGATCAGCCAGCAACACAG   447
332-2    34 ACATGGCTGTTGCCAAAGTATTAGCGTTTGTTAGTTTTGCAGTTTTTGTGCTAACGGTGTATGGCAGCAT-CGAT--GCTGATGAAATAG   120
            *    ***  *   *   **  ** ** *  *     * *  **         ** **  * ** *** *  ****  **     * * **

332-1   448 CAGATAAGAGAGCAAAATTGACAGATCTGAAGCTGCGACAACGTGAAGCAGACAATGCTATATTGGAATATGAACTTTCAGAGGCTAAAA   537
332-2   121 AATCATTAGAAAGAGAATTGACAGATCTGAAGCTGCGACAACGTGAAGCAGACAATGCTATATTGGAATATGAACTTTCAGAGGCTAAAA   210
             *     *  * ************************************************************************

332-1   538 GAGCAATAGATGCTTCATGCAATGACCAGCTAGGCAAGTCGAGATGTCAAAAGTACAGGAAATATGGATTTTGCAGAAAGGATTACAGAT   627
332-2   211 GAGCAATAGATGCTTCATGCAATGACCAGCTAGGCAAGTCGAGATGTCAAAAGTACAGGAAATATGGATTTTGCAGAAAGGATTACAGAT   300
            ****************************************************************************************

332-1   628 TAAAAAAGCTATGCAGAAAAACTTGTGGATTTTGTGGAGTAATGCCCAAAGTTCCACATTGTGCAAAGACTGCACTTGGATGCTGCTGGG   717
332-2   301 TAAAAAAACTATGCAGAAAAACTTGTGGATTTTGTGGAGTAATGCCCAAAGTTCCACATTGTGCAAAGACTGCACTTGGATGCTGCTGGG   390
            ******* ****************************************************************************

332-1   718 ATTTTCAGACTCCCAAGAAGGATGGAGCAGGAACTAACTGTCCAAAATGCAGAGACAATCCGAAAAAAAGAAGAGTTTGCAAAATGTTTG   807
332-2   391 ATTTTCAGACTCCCAAGAAGGATGGAGCAGGAACTAACTGTCCAAAATGCAGAGACAATCCGAAAAAAAGAAGAGTTTGCAAAATGTTTG   480
            ****************************************************************************************

332-1   808 AGCCTGACTGCAACAGTAACAAAGACGCTGGCAGCTTTATGAGGAAAACATGTCCGAGAACTTGTGGAGTATGTGGAGAAGGAGCTATGT   897
332-2   481 AGCCTGACTGCAACAGTAACAAAGACGCTGGCAGCTTTATGAGGAAAACATGTCCGAGAACTTGTGGAGTATGTGGAGAAGGAGCTATGT   570
            ****************************************************************************************

332-1   898 GCATGGATGATCCAGCAAAGGAAATGTACTGTGAAGAATGGAGTAACGAGGGAATGTGTGAAACTGAAAAACCAATGATGTCTGTTTATT   987
332-2   571 GCATGGATGATCCAGCAAAGGAAATGTACTGTGAAGAATGGAGTAACGAGGGAATGTGTGAAACTGAAAAACCAATGATGTCTGTTTATT   660
            ****************************************************************************************

332-1   988 GCAGGAAAACTTGCGGTATTTGCTAAGAAGTTGCCCATTATGCAAAAACACTTTCACGTTATTAGAAATTTTCACGCATTTTGCATTTTG  1077
332-2   661 GCAGGAAAACTTGCGGTATTTGCTAAGAAGTTGCCCATTATGCAAAAACACTTTCACGTTATTAGAAATTTTCACGCATTTTGCATTTTG   750
            ****************************************************************************************

332-1  1078 GAAACACCTTGTATTGGTCTTTTAGTAACTAGGAGTTTCTATTATTTCACGTTTGTAATATTATTAAGTTAATTTATGTAGAAAGAGTAA  1167
332-2   751 GAAACACCTTGTATTGGTCTTTTAGTAACTAGGAGTTTCTATTATTTCACGTTTGTAATATTATTAAGTTAATTTATGTAGAAAGAGTAA   840
            ****************************************************************************************

332-1  1168 TACTGGTTAAAGCACTTTCTTGACGTTAAGTATACTGTAAAGAAGATAGGCCTTTTTTTATTTTCATGGTATTGCCGGTATCAAAAAAAA  1257
332-2   841 TACTGGTTAAAGCACTTTCTTGACGTTAAGTATACTGTAAAGAAGATAGGCCTTTTTTTATTTTCATGGTATTGCCGGTATCAAATTTAA   930
            ****************************************************************************************   **

332-1  1258 AAAAAAAAAA---------------- 1267
332-2   931 AAAAAAAAAAAAAAAAAAAAAAAAA  957
            **********
```

Figure 5. 7 Nucleotide alignment of the Mk-332 isoforms.

The overall protein architecture of CbTX and Mk-332 isoforms is given in Figure 5.8 A and B, indicating that CbTX and Mk-332 are homologous proteins. The relative locations of the domains as predicted by SMART db (Schultz *et al.* 1998) are indicated in the left hand tables of Figure 5.8 A and B and in the accompanying diagrams. Confidence levels for the assignment are also indicated. The CbTX and Mk-332 consensus SXC/ShK motifs were aligned (Figure 5.8 C). Note that the first four residues have a distinctive X**C**X**D** pattern and invariably in each domain, three out of the six cysteines are found in the last eight residues with the signature **C**XX**TC**XX**C.** It is clear that the first four and the last eight residues have a low evolutionary rate in comparison with the central region. It was interesting to compare the box jellyfish SXC/ShK domains with the presumed homologous neurotoxic peptides from other species of cnidarians (Figure 5.8 D). A similar pattern of X**C**X**D** followed by **C**XX**TC**XX**C** is also seen in toxins from anemones and the antimicrobial "Aurelin". Note that the catalytic dyad (KY, KV or KL) identified in anemones (Gasparini *et al.* 2004; Gilquin *et al.* 2005) and "Aurelin" from a true jellyfish representative (Ovchinnikova *et al.* 2006) seems to be duplicated and changed position in some of the cubozoan ShK domains (compare Figure 5.8 C and D).

It is worthwhile to remark that initial BLASTP searches of CbTX-I/II and Mk-332-I/II sequences retrieved high e.values of sequence similarity with *Cns161* metalloproteinase 2 and astacin 3 from *Hydra vulgaris* (e.values of 1e-05 and 4e-05, respectively). The sequence identities are restricted to the fragment of the ShK toxin domain rather than the peptidase domain. In fact, searches at the MEROPS and SMART databases revealed that the amino acid residues identified as essential for catalytic activity in metalloproteases[7] are completely lacking in Mk-332-I/II and CbTX-I/II. Hence, these cubozoan gene products do not belong to the metalloproteinase family and it is uncertain whether these proteins lost metalloproteinase activity or whether they have arisen from duplication of a ShK-like gene from a basal anthozoan.

[7] The catalytic sites are usually HEXXH or more stringently defined as abXHEbbHbc; where a is most often valine or threonine, b is an uncharged residue and c an hydrophobic residue -Rawlings, N. D. & Barrett, A. J. 1995 Evolutionary families of metallopeptidases. *Methods in Enzymology* **248**, 183-228.

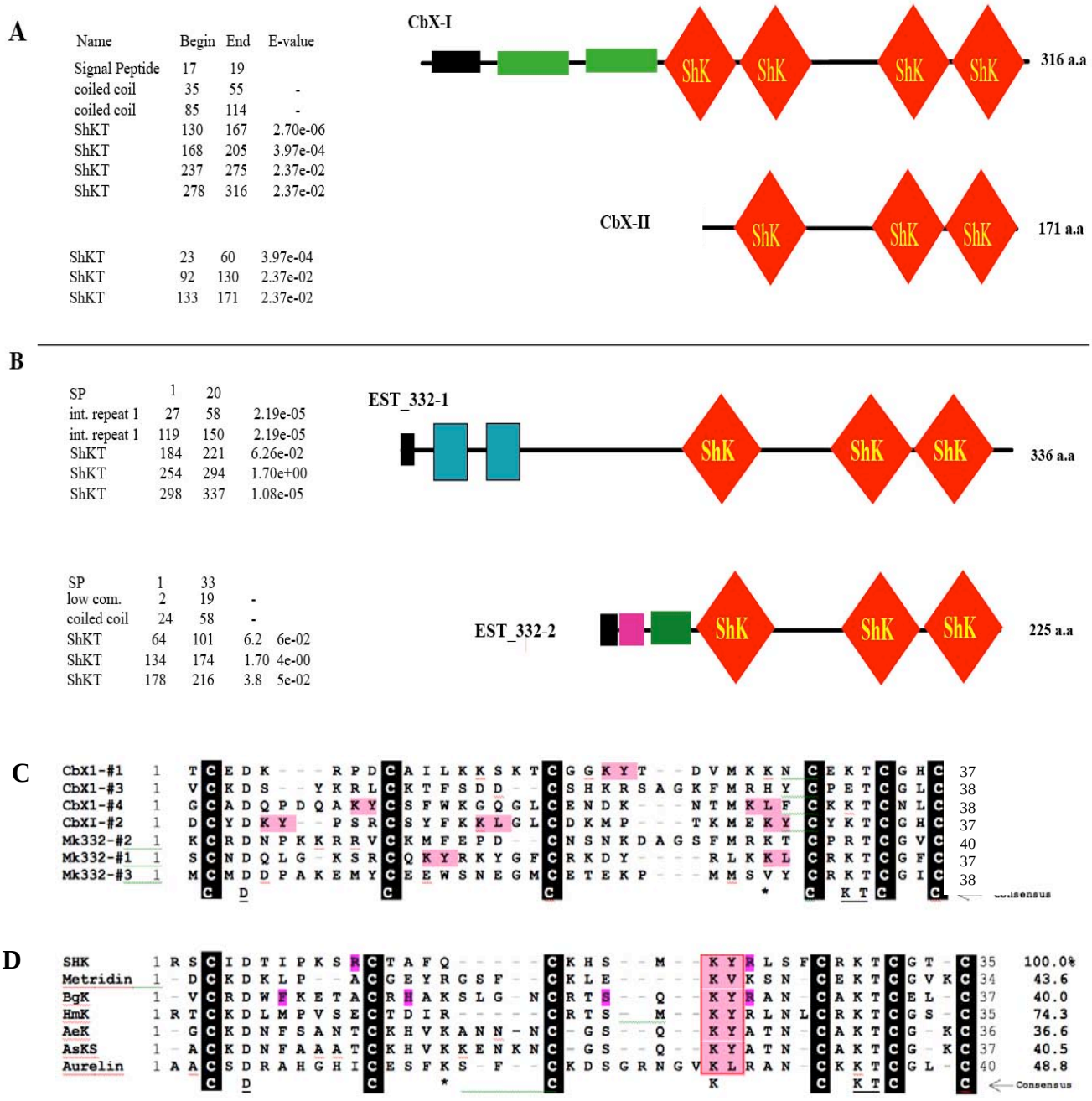


Figure 5. 8 Architecture of *C. barnesi* and *M. kingi* ShK domain-containing proteins.

Protein architecture of (A) CbX and (B) Mk-332 isoforms. (C) Alignment of the cubozoan ShK motifs, showing sequence similarity and conserved cysteines. (D) Sequence homology between sea anemone toxins and Aurelin, an antimicrobial with both defensin and possibly potassium channel blocker activities. Anemone toxins include ShK, a *S. helianthus* potassium channel toxin, Metridin from *M. senile*, BgK, a B. granulifera potassium channel toxin, AsKS (also called kaliseptine), an *A. sulcata* potassium channel toxin, HmK from *H. magnifica*, AeK from *Actinia equina* and Aurelin from the scyphomedusa *A. aurita*. Sequences were aligned at cysteine residues. The functional dyad is indicated in pink and boxed in red (a critical lysine and an aromatic amino acid). Other residues that are involved in interactions with potassium channels are indicated in magenta. Conserved amino acids are underlined.

If short chain neurotoxins are part of cubozoan nematocyst proteins, they are expected to be highly expressed. However both DNA hybridization and EST analyses failed to identify short chain K^+ neurotoxin-encoding transcripts in cubozoans. These results indicate that cubozoans lack the secreted short neurotoxin homologs to ShK from *S. helianthus*. Similarly, analyses of large data sets of genomic information from two model cnidarians (*H. magnipapillata* and *N. vectensis*) have shown an absence of short K^+ neurotoxins homologs. Interestingly, in both *H. magnipapillata* and *N. vectensis*, the ShK motif was identified in a tandem arrangement, and inserted in other genes such as astacin metalloproteases (Moran *et al.* 2008a; Sher *et al.* 2005b). The metridin-like toxin domain was also found either at the N- or C-terminus of chymotrypsin serine proteases from cubozoan (*M. kingi)*, scyphozoan (*A. aurita*) and anthozoan *(N. vectensis)* species (consult Chapter 4). ShK was also identified in a dermatopondin-toxic like gene and a pathogenesis-related gene from *M. kingi* (ABZ10815).

Despite the fact that orthologs of ShK short chain neurotoxins affecting potassium channel conductance were not identified, transcripts encoding preproproteins and a non-secreted protein, which contained an ShK-encoding domain, were identified and subjected to molecular characterization (e.g. Figure 5.8). Searches of several protein databases revealed that similar genes are found in *N. vectensis*[8], *Caenorhabditis elegans*[9], other parasitic animals and in amphioxus[10]. The proteins encoded by these genes are in several cases secreted proteins that are synthesized as propeptides, resembling the cubozoan genes described in this chapter. All of these gene sequences are available at non-redundant databases, although they remain uncharacterized and are not reported in the literature. However, in a study identifying putative pathogen recognition and antimicrobial genes from the nematode *C. elegans*, genes encoding secreted proteins that contain ShK domains were up-regulated when infected with *M. nematophilum*, a gram positive bacterial pathogen. The genes containing the ShK domain seemed necessary for the swelling response, implying an antimicrobial response role (O'Rourke *et al.* 2006). In addition, during the course of this research, the genome of another parasitic organism, *Brugia malayi*, became available and many genes containing the ShK motif were identified, which are likely to share a function with similar proteins found in other parasitic organisms such as *C. elegans.* In other research involving the plant parasitic

[8] SB_8208, SB_35845
[9] pharyngeal gland toxin-related protein 1 and 2 (ref|NP_491508.2, ref|NP_491509.1)
[10] EEA63348

nematodes *Globodera pallida* and *Globodera mexicana*, similar genes were identified and named *IVg9* and *IA7*. These display high similarity with the cubozoan CbTX-I/Mk-332 genes. The authors determined the genomic structure of *IVg9* and *IA7* and found that the sequence encoding the mature product did not contain introns. *In situ* hybridization studies showed that *IVg9* and *IA7* are expressed in high levels in salivary glands and in the intestine and are assumed to be involved in pathogenicity (Blanchart *et al.* 2007).

It is attractive to suggest that the diversification of the ShK domain may be enhanced by the lack of introns in the gene, in conjunction with alternative splicing of mRNA and RNA editing events. This is a feasible scenario because alternative splicing has been a major factor in the functional complexity of genomes. For instance, it is estimated that 40-70 % of human genes have alternatively spliced transcripts (Namshin *et al.* 2005), and alternative splicing has been the main strategy for expanding venom diversity, for example, in scorpion (Zhijian 2006) and snake venoms (Pung *et al.* 2006).

Intriguingly, the ShK-like toxin domain is also found in pathogenesis-related genes from monocotyledon plants and also from humans, and at least in humans it is suggested that the domain may contribute to innate immunity (Kitajima & Sato 1999). Importantly, in this research a similar pathogenesis-related gene was found in *M. kingi* (GenBank accession number ABZ10815).

5.4.3 Spatial distribution of CbTX-I mRNA signal

To identify the sites of expression of CbTX-I mRNA and to infer a possible function, whole mount *in situ* hybridization was conducted in a juvenile *C. barnesi*. A single-stranded RNA probe was generated by digoxigenin-labelling of the the full length CbTX-I cDNA. The results (see Figure 5.9) show differential mRNA expression, observed in three major areas:

i) The stomach pouch manubrium and in single secretory cells in the mouth area

ii) In the tentacle battery of nematocyte cells. Since nematocysts are cellular organelles and not not cells, no positive signal was detected within them, and instead positive signal for CbTX-I is attributed to the underlying nematocytes, responsible for nematocyst biogenesis

iii) In the distal/internal region of the pedalia.

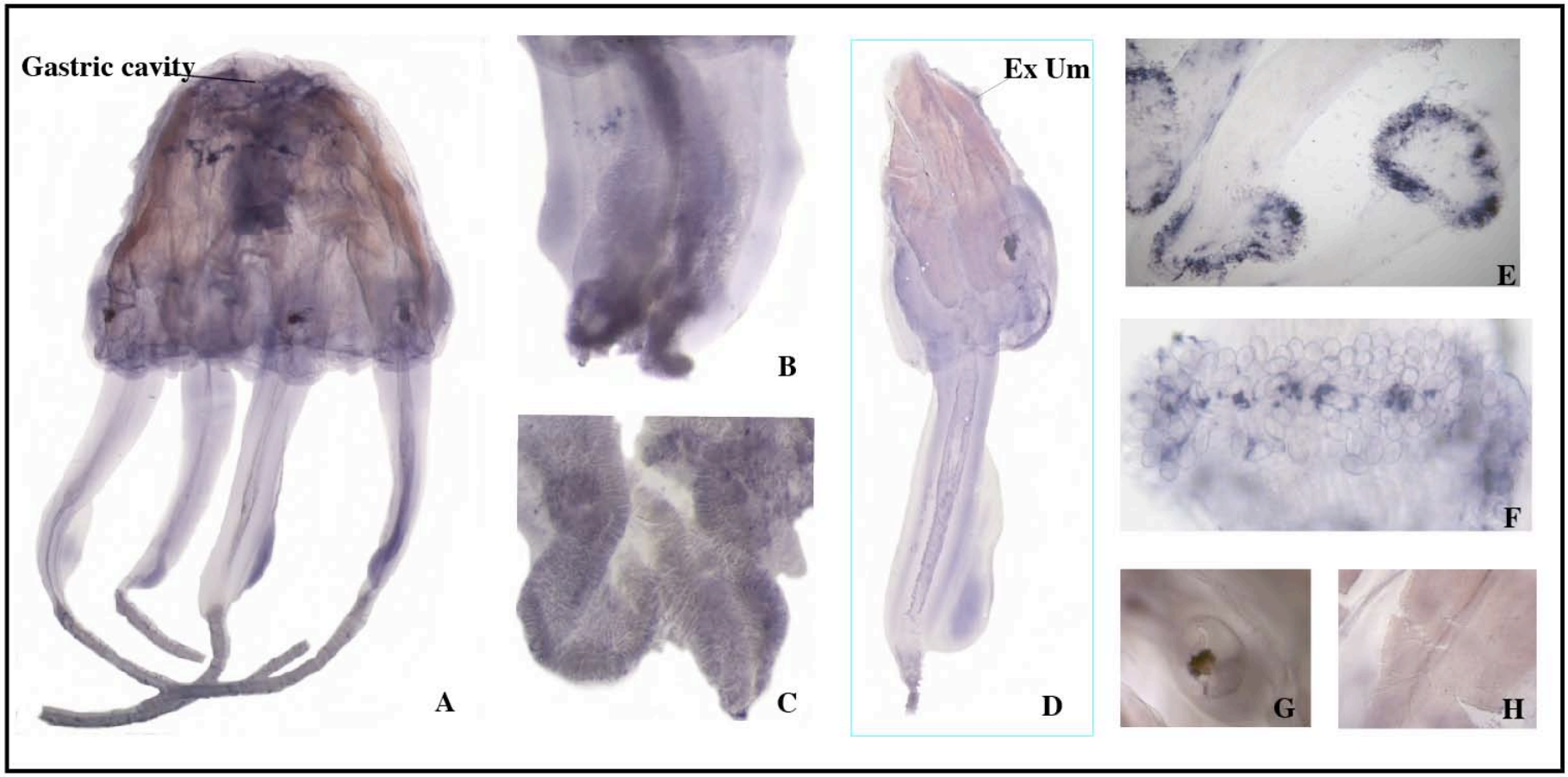


Figure 5. 9 CbTX-I transcript localization by whole mount *in situ* hybridization of *C. barnesi.*

(A) Juvenile *C. barnesi* whole specimen indicating the expression of CbTX-I using digoxigenin-labelled RNA probe. (B and C) Ectodermal expression in the gastrodermic tubule (manubrium) and in epithelial cells at the lips. (D) Disparate ectodermal expression at the pedalia area and margen sub-umbrella pattern. Gene expression in the cells located at the base of the nematocyst battery. (G and H) For the purpose of comparison, note regions with no mRNA signal like an eye and a gonad fragment.

Since CbTX-I and CbTX-II cDNAs are very similar, it is assumed that the mRNA signal detected corresponds to both genes. Expression of CbTX-I in the stomach, mouth and nematocytes is likely to be related to defence mechanisms. In contrast, the signal in the pedalia may be due to cross-hybridization with other genes containing the ShK domain. For example, some of the astacin metalloproteases in cnidarians (as well as in other invertebrates) contain the toxic motif. *In situ* hybridization studies in *Hydra* reveal the involvement of astacins in pedalia morphogenesis (HMP-1), head regeneration and also in trans-differentiation of tentacle battery cells (Pan *et al.* 1998; Yan *et al.* 2000).

5.4.4 Spatial distribution of Mk-332-I mRNA signal

According to the primary structure of Mk-332-I/II, both transcripts encode basic polypeptides comprised of three ShK toxic domains with extensive cysteine linking (Figure 5.8). It was relevant to compare the gene expression pattern of Mk-332 in juvenile and adult Irukandji specimens in order to determine whether these genes are involved in defence mechanisms and are equally expressed through different stages of the life cycle. The results of this experiment (Figure 5.10) illustrate that there are differences and similarities in the expression pattern of Mk-332 in juveniles and adults. In the juvenile Irukandji, expression of Mk-332 was detected in the gastrodermis, lip epithelial cells, in the manubrium and in clusters of nematocytes cell in the tentacles (Figure 5.10 B-D and F). In contrast, expression of Mk-332 was only detected in the tentacular nematocytes of the adult jellyfish (Figure 5.10 I and J, F and H).

It is relevant to point out that the specimens used in this experiment were *C. barnesi* due to the limited availability of *M. kingi* specimens. In the last decade no more than 15 specimens of *M. kingi* have been collected, and only three specimens were available for this research (including those used to construct a cDNA library). However, as mentioned above, CbTX-I/II and Mk-332-I/II DNA sequences and protein domain architectures point to gene homology and common protein function. Accordingly, the use of the *M. kingi* 332-II riboprobe for *C. barnesi* whole mount *in situ* hybridization experiments resulted in cross-hybridization among orthologous genes. In fact it was a useful strategy to target the cubozoan ShK toxin-like genes and, at the same time, reduce background signal that may be generated by transcripts of non-homologous genes containing the conserved ShK domain.

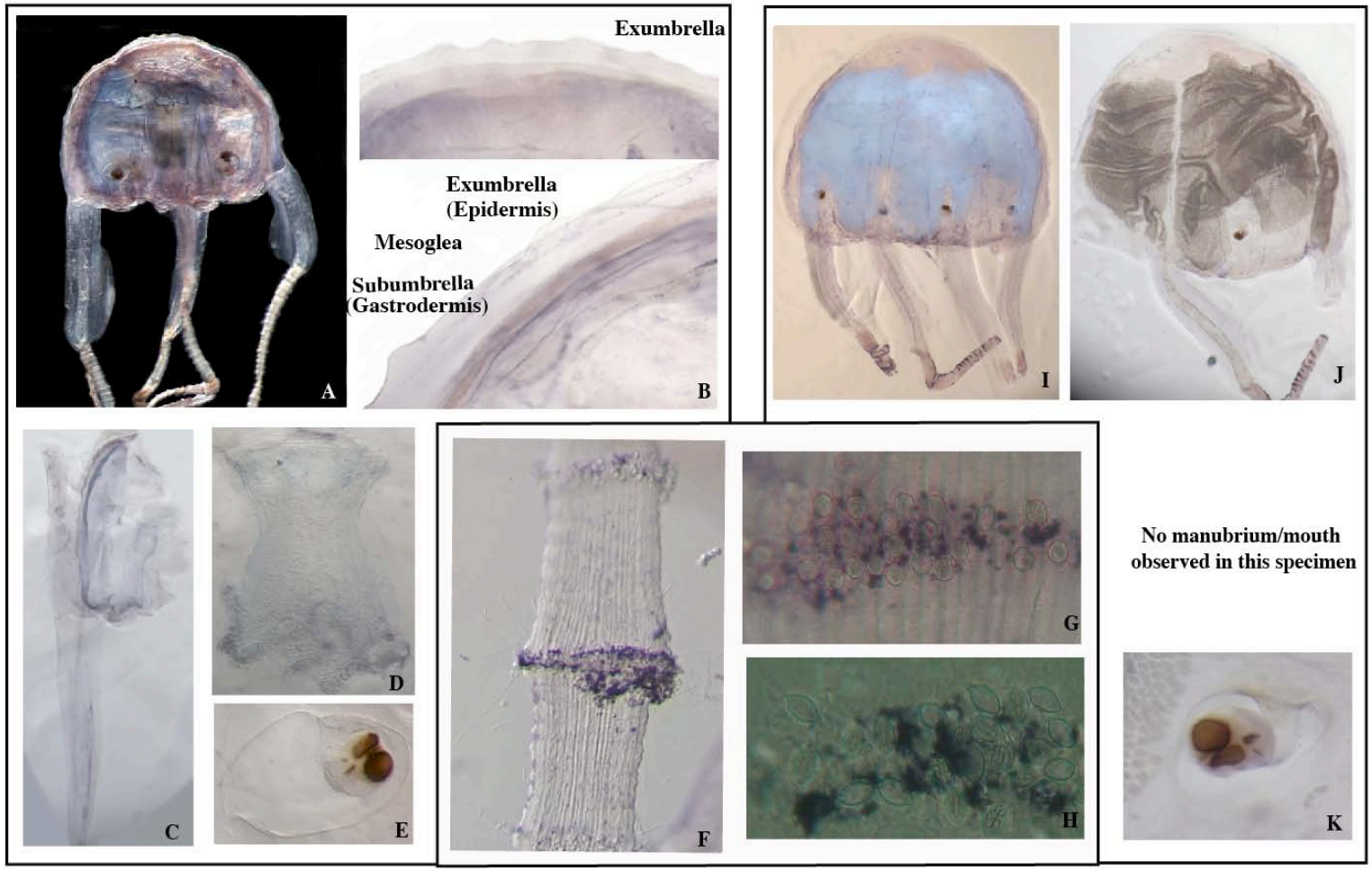


Figure 5.10 mRNA expression pattern of Mk-332-I in juvenile and adult *C. barnesi* specimens.
(A) Complete Irukandji juvenile; (B), (C) and (D) correspond to vertical cuts in close up view of umbrella, pedalia and manubrium, respectively. (E) shows the eyes, (I) and (J) show a reproductively mature specimen and a vertical cut view of the entire Irukandji jellyfish. (F to H) show mRNA signal located in tentacles, both in juveniles and adults. Note the stained cluster of nematocyte cells.

A comparison of Figure 5.9 (A to C) and Figure 5.10 (A to D) shows that the expression patterns observed when using the CbTX-I/II and MK-332-I/II probes for juvenile *C. barnesi* are very similar (endodermal cells in the gastrodermis, manubrium and in epithelial cells at the lips, and in tentacles where expression is ectodermal in nematocytes). The exception is that when the Mk-332-I/II probe was used, no expression was detected in the pedalia or in the exumbrella. These results indicate that expression of CbTX-I/II in the pedalia was due to non-specific cross hybridization. In the adult specimen of *C. barnesi* probed with Mk-332-I/II, the expression was different compared to the juvenile specimen treated with the same riboprobe indicating gene expression changes between two ontogenic stages. The differences of gene expression between two stages of life may be due to changes in diet and sexual maturity. In figure 5.10 I and J shows an specimens with fully developed gonads, it may be that at that stage the endodermal area are more engaged in genes related to reproduction rather than genes devoted for depredation such as the proteins reported in this work.

5.4.5 Bacterial over-expression of box jellyfish proteins

CbTX-I and II were sub-cloned into a pQE expression vector system for bacterial over-expression. The proteins were expressed with a leader sequence and the propeptide regions to enhance expression. Expression of the propeptide can assist precursor trafficking and these forms of the protein lack putative enzyme activity. This was an important factor given that these specific proteins may have cytotoxic activity. However, it has been possible to obtain fully active classical cnidarian short chain neurotoxins affecting potassium conductance (HmK from *Heteractis magnifica*) in a bacterial host without the propeptide part of the protein (Gendeh *et al.* 1997b).

Small-scale expression and purification of CbTX-I and CbTX-II were conducted. Coomassie-blue staining after protein separation revealed protein bands at approximately 21 and 35 kDa, the sizes expected for CbTX-I and II recombinant proteins (Figure 5.11). Western blot detection of protein bands that were antigenic to anti-tetra-histidine antibodies confirmed that CbTX-I and II were being expressed.

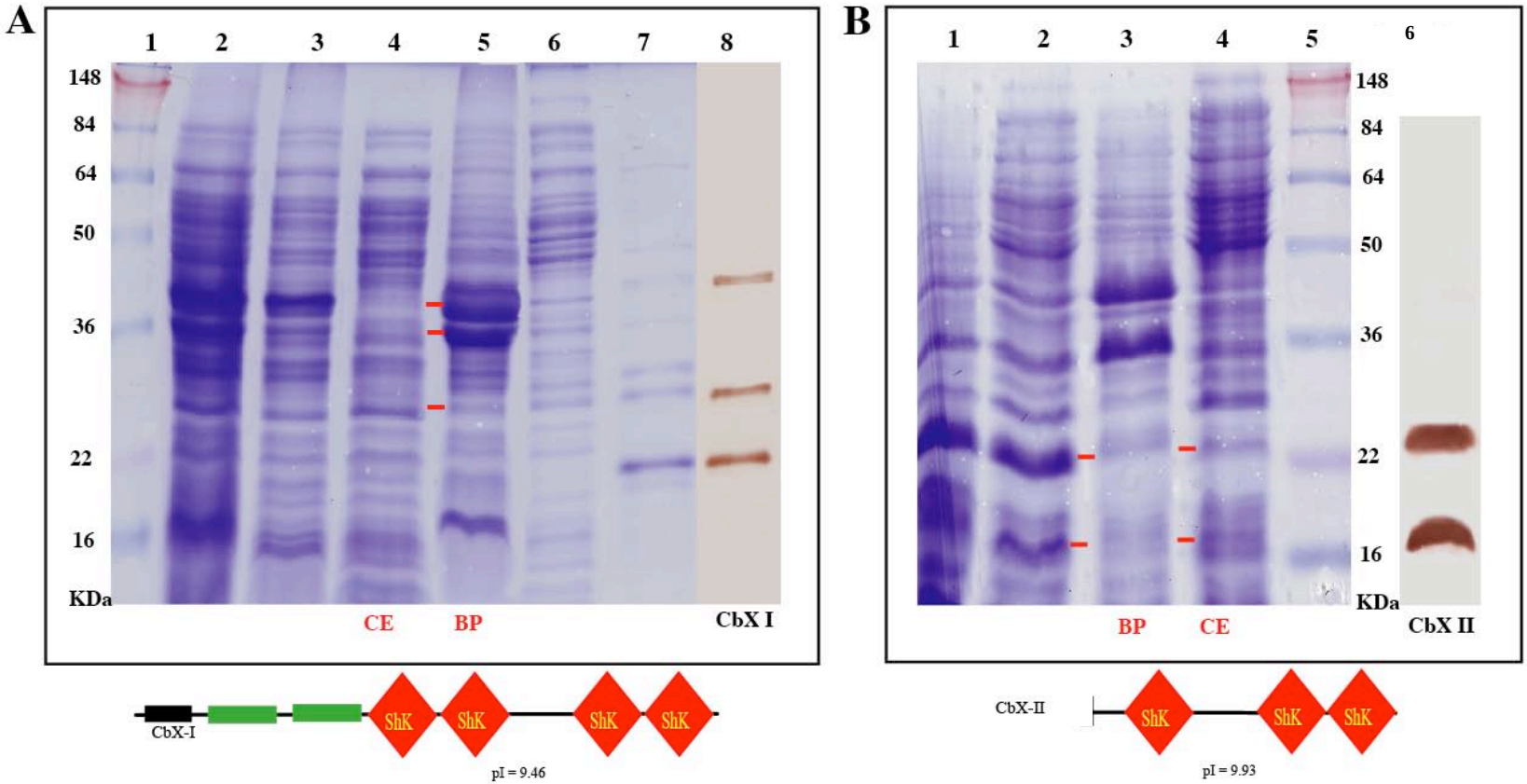


Figure 5. 10 Recombinant protein expression of CbTX-I and CbTX-II.

(A) Lane 1 - molecular weight markers; lane 2 - bacterial suspension; lane 3 - sonicated bacterial suspension; lane 4 - crude extract; lane 5 - sonicated bacterial pellet; lane 6 - column flow through; lane 7 - protein eluted with 50 mM imidazole; lane 8 - western blot of His-tagged protein. (B) Lane 1 - bacterial suspension; lane 2 - sonicated bacterial suspension; lane 3 - bacterial pellet; lane 4 - crude extract; lane 5 - molecular weight markers; lane 6 - western blot of His-tagged protein.

Preliminary experiments with CbTX-I recombinant protein proved it to be lethal in insects producing classic neurotoxic effects: tetanic movements followed by prolonged paralysis and death, and similar results have been obtained with recombinant and native venom proteins from invertebrates (Martin-Eauclaire *et al.* 1994; Shu & Liang 1999). It is anticipated that Mk-332-I/II may exhibit neurotoxic activity, however not enough recombinant protein was obtained for further experiments and optimization of recombinant production is required.

5.5 CONCLUSIONS

Remarkably, despite an intensive search of the EST library collection and thorough screening of a cDNA library by DNA hybridization, orthologs to the classical short chain neurotoxins affecting potassium ion conductance (similar to those in *S. helianthus*) were not found. Instead CbTX-I/II and Mk-332-I/II were discovered. These cubozoan genes display a tandem arrangement of motifs that are extensively cysteine cross-linked and strongly resemble the potassium channel blocking toxins from anthozoan mini genes (ShK, Hmk, BgK etc.). Apart from the ShK toxic domain, the presence of any other protein domains was not distinguished in CbTX-I/II and Mk-332-I/II.

Based on the overall DNA sequence similarity and domain architecture, CbTX and Mk-332 (and their respective isoforms) are orthologous genes from the Irukandji box jellyfish *C. barnesi* and *M. kingi*.

Comparisons of the cDNA sequence indicate that the CbTX transcript isoforms have several key differences. CbTX isoforms I and II are dissimilar at the 5′-ends, including nucleotide insertions and deletions. A deletion of only two nucleotides in the coding region led to a frame shift producing a major change in the translated products, a preproprotein (CbTX-I) and a non-secreted protein (CbTX-II). At the 3′-end a major nucleotide insertion in CbTX-I in comparison to CbTX-II was located between the last of five polyadenylation signals and the poly A tail. Jointly, these observations are clear indications that alternative RNA editing events have taken

place in the origin of these *C. barnesi* transcripts. Despite these differences, both CbTX isoform transcripts encode exactly the same mature polypeptide.

Nucleotide differences at the 5′-end between Mk-332-I and II variant transcripts were also found and are more substantial than those found in the *C. barnesi* counterparts. For instance, 120 bp upstream of the AUG start codon and before the signal peptide-encoding region of Mk-332-I there was an insertion of 325 bp containing two close DNA repetitions. Moreover, because of this insertion the two isoforms encode different signal peptidase recognition sites. Mk-332-II has only one predicted proconvertase-binding site that is identical to the last of the four possible propeptide cleavage regions identified in Mk-332-I. However, apart from the transcript and the prepro region differences, the mature polypeptide has homology to the CbTX isoforms.

At the N-terminus of the mature proteins, CbTX-I and Mk-332-I and -II have signal peptidase recognition sites and a long propeptide region ending in polar and charged residues. Similar dibasic cleavage sites have been reported, before mature nematocyst and non-nematocyst secreted toxins from other cnidarians (Anderluh *et al.* 2000b; Hasegawa *et al.* 2006; Honma *et al.* 2005a; Moran & Gurevitz 2006; Ovchinnikova *et al.* 2006), indicating that these cubozoan gene products could follow similar maturation pathways.

In regards to protein expression, CbTX isoforms I and II were bacterially expressed and yielded several protein bands as detected by SDS-PAGE and western blot analyses, indicating protein cleavage in agreement with the predicted analysis of the translated sequences. It is unclear, however, whether all of these cleavage fragments are biologically active peptides or whether some may represent artificially degraded protein fragments. Neurotoxic biological activity is anticipated on the basis of preliminary experiments using bacterially expressed CbTX-I and -II. It could be the case that the motifs arranged in tandem may act synergistically but this hypothesis has to be evaluated experimentally.

The spatial distribution of CbTX-I and Mk-332-II mRNA signals in juvenile *C. barnesi* specimens were similar and indicated that these genes are highly expressed in discrete regions specialized for digestion and prey capture. It is relevant that some differences were observed. Hybridization at the margin of the sub-umbrella and pedalia was not observed when the *M. kingi* probe was used on *C. barnesi* specimens. This could indicate non-specific hybridization with an evolutionary related gene. For instance gene products containing the ShK domain are involved in development (particularly in pedalia morphogenesis) in hydrozoans. Importantly, differential expression of Mk-332-II between juvenile and adult specimens was detected and revealed that in adults, gene expression was restricted to the battery of nematocytes.

CbTX and Mk-332 isoforms are regarded as toxic proteins, and members of the ShK toxin family due to i) similarities of the prepropeptide regions with other cnidarian toxins, ii) similarity of the ShK conserved domain and the tandem arrangement, iii) the gene expression patterns detected by *in situ* hybridization and iv) the neurotoxic activity induced in insects using the CbTX-I and -II recombinant proteins.

Box jellyfish are agile, voracious and aggressive creatures that fast recover their toxic gadgets.(nematocytes) The overall gene and domain structure of CbTX and Mk-332 isoforms may reflect the predatory nature of box jellyfish, in which toxic genes are expected to quickly evolve for fast response, efficient synthesis, maximum damage and quick self recovery.

5.6 FUTURE DIRECTIONS

Genomic DNA data from cubozoans is very poor and, to date, no DNA genomic libraries exist (e.g Cosmid, BAC libraries etc). Nonetheless, this molecular resource is required and when available would be very useful in many respects. For instance, in this research chapter the characterization of the cDNAs and their translated preproproteins as well as the non-secreted protein (CbTX-I) leaves an open door to further investigate whether the mRNA subtypes of Mk-332 and CbTX are products of independent genes or derived from alternate splicing and editing of coding regions of RNA. Therefore, the availability of genomic sequence data would allow greater

insight into the splicing regions (intron/exon boundaries), as well as assist with more accurate identification of promoter regions and basal transcriptional regulator elements including TATA and CAAT boxes. Genomic data may also provide evidence to reconstruct the evolutionary origin and diversity of the ShK encoding genes among cnidarians and other invertebrate phyla in which similar protein architecture is seen.

It is not straightforward to infer the origin and diversification of the ShK/SXC/Tox1 domain. A comprehensive phylogeny has to be taken into consideration, including all proteins containing the domain. However, the domain has been recruited by many different proteins (in some cases containing more than two domains), and this makes it difficult to trace ShK evolutionary history.

It has been reported that nematocyst proteins from *C. barnesi* show variation during ontogeny (Underwood & Seymour 2007), however these results could benefit by whole mount and cryosection *in situ* mRNA hybridization of cubozoan defence genes.

With regard to the expression studies, overproduction of recombinant proteins was achieved and the expressed protein could be used to investigate the toxic properties of these proteins and to further characterize how these proteins are involved in the defence system of box jellyfish. In addition, CD spectrometry and/or fluorometry studies may assist in elucidating the structure and disulphide cross-linking pattern among the ShK domains.

CHAPTER 6

Identification and molecular characterization of two putative cytotoxic proteins from the Irukandji box jellyfish *M. kingi*

6.1 ABSTRACT

This chapter reports the identification and molecular characterization of two abundantly expressed genes that encode putative cytolytic toxins, MkTX-A and -B, from the potentially lethal Irukandji box jellyfish, *M. kingi.* The genes were identified by comparative analyses of a small, non-redundant expressed sequence tag (EST) collection and DNA hybridization screening of a *M. kingi* cDNA library. Molecular information available in public databases as well as primary sequence analyses indicated that MkTX-A and -B are related to a novel family of pore-forming cytotoxic proteins from box jellyfish, but also have significant sequence similarity with recently released sequence data of *Hydra magnipapillata* and distantly related salivary proteins from arthropods.

Analyses of primary and secondary structures indicated that the first 50 amino acid residues at the N-terminus of mature MkTX-A contains two small amphiphilic α-helical regions and, according to the high surface index score, it is predicted to occur on the protein surface. The water-soluble portion of the amphiphilic α-helices may interact with lipids in a covalent manner for initial attachment. Three transmembrane regions were also predicted, the first of which is long enough to span a biological membrane. The presence of this transmembrane spanning region suggests that MkTX-A may act via a pore-forming mechanism to disrupt normal transmembrane ion concentration gradients in susceptible cells**.**

In contrast to other cnidarian basic charged pore-forming peptides, such as well studied actinoporins (Anderluh & Lakey 2008; Kristan *et al.* 2007), MkTX-A, MkTX-B*f* and the five other members of the cytolytic box jellyfish toxin family (CrTXs, CaTX-A, CfTX-1/-2 and CqTX-A) were predicted to be cysteine cross-linked polypeptides with a protein scaffold comprising two structurally different domains:

the N-terminus is dominated by an α-helix bundle domain and a β-strand arrangement of approximately 150 amino acid residues is located at the C-terminus of the proteins. Similar structures were also predicted for two hypothetical hydrozoan homologs inferred from the genome of *H. magnipapillata*. The C-terminal domain of MkTX-B displayed remote structural similarity to the ricin β-like protein superfamily of lectins (SCOP b.42.2.1). Equivalent predictions for other carybdeid cytolysins (CrTXs and CaTX-A) were obtained, but not for the chirodropid cytolysins (CfTX-1/-2 and CqTX-A). Dissimilarities between the two structural motifs could indicate functional variation between Orders Carybdeida and Chirodropida.

Keywords: Irukandji; Box jellyfish; Cubozoa; ESTs; Cytolysins.

6.2 BACKGROUND

Class Cubozoa includes less than 30 known marine species, some rare and tiny, and others considered the most lethal animals on the planet. At the same time they are very primitive organisms. Despite this, these fascinating, ancient and agile creatures possess image-forming eyes, however unlike higher metazoans, they have no brain to process the information. Compared to other cnidarians, cubozoans not only possess an elaborate behaviour but also have evolved a complex and efficient repertoire of toxic components. Cubozoan venoms are well-documented to induce physiological and toxicological effects such as lethality, haemolysis, dermonecrosis as well as neurotoxic and cardiotoxic effects (Carrette & Seymour 2006; Nagase *et al.* 1986; Ozaki *et al.* 1986; Ramasamy *et al.* 2005a; Sanchez-Rodriguez *et al.* 2006; Winkel *et al.* 2003; Winkel *et al.* 2005; Winter *et al.* 2008). Some species of cubozoans, specifically of the Order Carybdeida, such as *Carukia shinju, Malo maxima* (Gershwin 2005b), *Alatina mordens* (Winter *et al.* 2008), *M. kingi* (Gershwin 2007a), Morbakka (Williamson et al 1996 or Tibballs 2006) and *C. barnesi* (Barnes 1964) are capable of producing Irukandji syndrome in humans who accidentally come into contact with the nematocysts of these creatures. Depending on the species and the degree of contact, the clinical symptoms vary in potency from mild to very severe stings. Although most Irukandji stings are not life threatening, two fatalities from

cardiac failure have been reported so far (Fenner & Hadok 2002; Fenner & Hadok 2003). *M. kingi* is believed to have been responsible for one of these deaths (Gershwin 2006b; Gershwin 2007b). Irukandji stings are commonly but not always slight with no immediate pain. There is a characteristic 30 minute delay in the onset of systemic symptoms, including agonizing and generalized pain, catecholamine-like effects such as arterial hypertension, distress, salivation, sweating, and gastrointestinal symptoms such as abdominal pain, nausea and vomiting (Fenner *et al.* 1986b; Huynh *et al.* 2003; Kinsey *et al.* 1988; Little & Mulcahy 1998). Pulmonary oedema has been observed and also cardiac dysfunction followed by death has occurred in two fatalities (Fenner & Hadok 2002).

In vivo and *in vitro* analyses of Irukandji venom have lead to the suggestion that the physiological effects of Irukandji stings are the result of elevated endogenous catecholamine release followed by cardiac collapse (Winkel *et al.* 2005; Winter *et al.* 2008). This suggests that Irukandji venom disrupts excitable membrane homeostasis, however the agent causing this effect has not been elucidated. In contrast, a small group of potently haemolytic and potentially lethal toxins has been identified and characterized from four species of cubozoans other than Irukandji (Brinkman & Burnell 2007; Nagai 2003). The nature of the interaction between cubozoan cytolysins and lipids in the membrane lipid bilayers and the role of the pore forming mechanism leading to leakage of internal contents suggested by Nagai (2003) and Brinkman & Burnell (2007) are not known. Therefore, the aim of this chapter was to search an EST library of the Irukandji jellyfish, *M. kingi*, to find genes encoding homologous toxins to those previously reported in other species of cubozoans.

6.3 EXPERIMENTAL PROCEDURES

6.3.1 Sample collection

Specimens of *M. kingi* were collected during the late summer months of 2004-2005 by Russell Hore from the outer Great Barrier Reef at Port Douglas, North Queensland, Australia. Samples were kept alive and once on land were snap-frozen and transported to the laboratory where they were stored at -80^{o}C until required.

6.3.2 cDNA library construction, mass excision

Total RNA was isolated from two mature whole specimens (530 mg) by disruption with glass beads in Trizol reagent (Invitrogen) according to the manufacturer's instructions. RNA samples were tested for integrity in a 1% (w/v) agarose gel. mRNA was isolated using polyATrack as described by the manufacturer (Promega). The cDNA library was generated using lambda ZAP-cDNA synthesis and Gigapack III Gold Cloning kit (Stratagene). The phage library was transformed into a phagemid library using the mass excision protocol of the supplier. Transformed *E. coli* were plated on agar containing carbenicillin and plates incubated overnight. cDNA clones were randomly picked. Plasmid DNA was purified (Qiagen) and both ends of the inserts were sequenced (Macrogen Inc. Korea).

6.3.3 Sequence identification and cDNA clones isolation

EST sequences from the *M. kingi* cDNA library were compared to GenBank nucleotide and protein databases using BLASTN, BLASTX and BLASTP programs at NCBI (Altschul *et al.* 1990; Gish & States 1993) in order to identify putative gene products involved in defence. EST_269, a 3'-end read, had a best match with gb|ABS30941.1| toxin-2 from *Chironex fleckeri* with and e. value of 3e-09.

6.3.4 Isolation of full-length cDNA clones encoding MkTX-A

cDNA clones were obtained by screening the *M. kingi* cDNA library using a gene-specific radiolabelled DNA probe. The probe was prepared using the SK polylinker restriction endonuclease sites *Apa*I at the 5′-end and *Pst*I at the 3′-end in conjunction

with an internal *Eco*RI site of the cDNA clone from EST_269. Plasmid DNA was digested with either a combination of *Apa*I and *Eco*RI or *Eco*RI and *Pst*I (New England Biolabs Inc). Digested DNA fragments were excised from an agarose gel and purified (QIAquick, Qiagen). Purified fragments (25 ng) were labelled with [α-^{32}P] ATP using a DECA Prime II kit (Ambion) and used to screen 5 x 10^6 *pfu* from the *M. kingi* cDNA library. Eighteen primary positive clones were picked. Insert sizes were determined by PCR using a combination of the forward SK universal T3 primer and a gene specific reverse primer (Gpr89) 5′-CACATACCAAGTACTTTGGTAGCTC-3′. PCR conditions were adjusted according to primer nucleotide composition as well as the expected length of the insert. The following profile was typically used: 95°C for 5 min, 32 cycles of 95ºC for 45 sec, 50ºC for 1 min and 72ºC for 2 min with a final extension temperature of 72ºC for 5 min. Six plaques from the *M. kingi* cDNA library were subjected to secondary screening with the radiolabelled probe and independent clones were isolated. ExAssist helper phage (Stratagene) was used to excise, *in vivo*, the pBluescript phagemid from the λ Uni-ZAP vector and finally plasmid DNA obtained by QIAprep (Qiagen) was sequenced in both directions.

6.3.5 Molecular characterization of cDNAs and translated products

The presence and location of signal peptide cleavage sites of precursor amino acid sequences were predicted using the SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen *et al.* 2004a). The predicted mature sequences were used to search for homologous proteins using the public BLAST platform at NCBI (Gish & States 1993), located at http://www.ncbi.nlm.nih.gov. Sequences that gave the best scores were compared using the alignment program ClustalW at the EBI web site (http://www.ebi.ac.uk/clustalw/). Alignments were edited using MacVector. The secondary structure was predicted using a combination of MacVector Toolbox and the GOR (Garnier Osguthorpe and Robson) IV algorithms. The TMpred program was used to predict membrane-spanning regions and their orientation (http://www.ch.embnet.org/software/TMPRED_form.html) (Hofmann & Stoffel 1993). Cysteine disulphide bonding states and connectivity predictions were done using DISULFIND (http://disulfind.dsi.unifi.it/monitor.php?query=5dkZ1z (Ceroni *et al.* 2006). Hypothetical three-dimensional structural models were generated by

SAM_TO8 (George Shackelford 2007; Katzman *et al.* 2008) or I-TASSER (Wu *et al.* 2007; Yang 2008). Model structures were studied with Swiss-PDBviewer http://www.expasy.org/spdbv (Guex & Peitsch 1997). Phylogenetic analyses of amino acid sequences were undertaken using CLS bioA/S sequence viewer version 6.02 (under default parameters) and MolPhy version 2.3 using the Dayhoff substitution Matrix and local rearrangement search mode (Adachi & Hasegawa 1996).

6.4 RESULTS AND DISCUSSION

6.4.2 Isolation and molecular characterization of MkTX-A and MkTX-B

EST_269 was identified by comparative analyses of an EST collection using the BLASTX algorithm (Altschul *et al.* 1990). The program detected significant homology of EST_269 with four sequences at the gene bank: corresponding in order of similarity to CfTX-2 from *Chironex fleckeri* with an e. value of 3e-09), CqTX-A from *C. quadrigatus* (3e-07), CfTX-1 (4e-07) and CrTX-A/CrTX-B Toxin precursor with a confidence value of 0.30. (Accession numbers: ABS30941.1, P58762, ABS30940.1 and Q9GV72).

The similarity of EST_269 to box jellyfish toxins lead to name the mature protein MkTX-A. The two most complete cDNA clones encoding MKTX-A (isoform 1 and 2) were obtained by DNA hybridisation of the *M. kingi* cDNA library. In addition it identified "MkTX-B*f*" a non complete but nevertheless related gene to MkTX-A.

Two isoforms of MkTX-A were discovered in the cDNA library. MkTX-A isoform I and 2 are 1706 and 1703 bp in length with open reading frames of 1374 and 1362 bp, respectively. Both MkTX-A isoforms lack the initiating methionine. The proteins encoded differ only by an insertion of three additional, consecutive amino acids (alanine, aspartate and alanine) in the putative signal peptide of isoform 1 (see Fig. 6.1). The mature protein is exactly the same and it is 434 amino acids long. The sequences have been assigned NCBI Accession numbers: *M kingi* cytotoxin A (MkTX-A) isoform 1 FJ705827, MkCTX-A isoform 2 FJ705828. The molecular mass, amino acid composition, isoelectric point and aliphatic index of mature MkTX-A correlate well with all other related cubozoan toxins (See Figure 6.4). Mature MkTX-A is basic and has a calculated mass of 48.55 kDa.

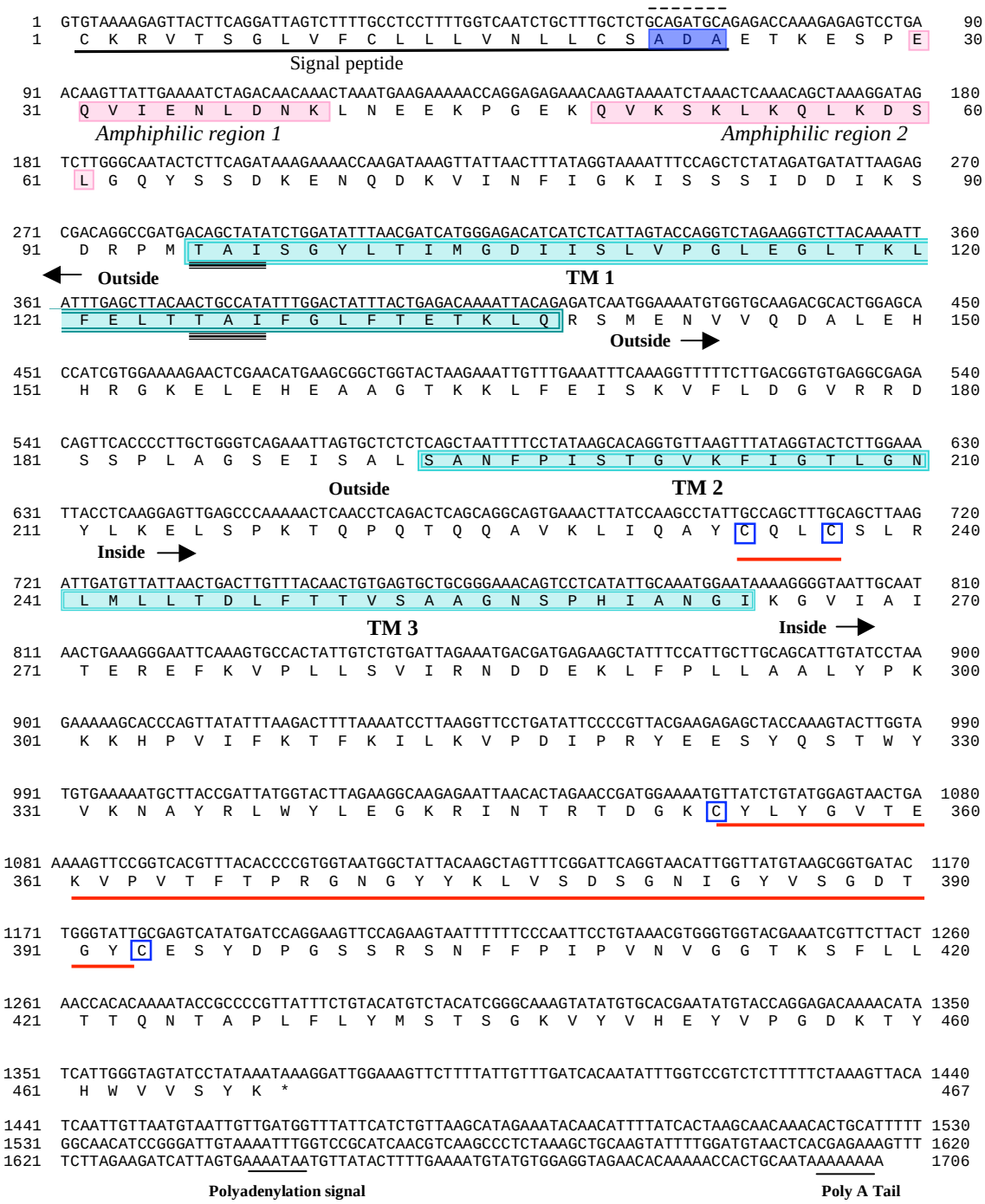


Figure 6. 1 Nucleotide and deduced amino acid sequence of MkTX-A isoform 1.

Amino acids underlined in black indicate the signal peptide. Cysteine residues are boxed in blue. Double underlining indicates a repetitive region (threonine, alanine and isoleucine). Predicted disulphide bridges are indicated in red. Putative amphiphilic α-helices and transmembrane-spanning (Hayward *et al.*) regions are highlighted in pink and aqua, respectively. The orientation of each TM region is indicated by arrows. A 3-residue sequence variation between MkTX-A isoforms 1 and 2 is highlighted in purple.

MkTX-B clone was truncated and several attempts to obtain the full length cDNA sequence by PCR using purified λ DNA were unsuccessful. The truncated cDNA clone was 922 bp (NCBI accession number: FJ705829), Figure 6.2. The first 434 bp translated into a product of 143 amino acids displaying high e values with the C-terminal portion of CrTX-A and CaTX-A (3e-46 and 9e-41, respectively), and to a lesser extent, similar to MkTX-A and the chirodropid homologs CqTX-A, CfTX-1 and -2 (similarity with the last three was not detected by BLASTP). The fragment contains three cysteines in the translated sequence, the second of which is conserved in all other members of this protein family.

```
  1 ATTTGAAGCATTTTGGTGTTCCTGACCCACCTTCAAGCTTTAAAAGTTCAAGTTACTCATTTAGAAATATCTATTGGAAATCCTGGTCTA   90
  1    L  K  H  F  G  V  P  D  P  P  S  S  F  K  S  S  S  Y  S  F  R  N  I  Y  W  K  S  W  S    29

 91  TTTGTGTTGAGAGTTACTTACAATATTACATCTTCAGAGGCTGTCCGAAACGTCTGCCCCATTATGTAAATGTAAACAAACTTAGCACTG 180
 30  I  C  V  E  S  Y  L  Q  Y  Y  I  F  R  G  C  P  K  R  L  P  H  Y  V  N  V  N  K  L  S  T   59

181  GTCATTACACAATGGAACTAAGTAAGAGAACGCTCTATGTAACAAAGCATGAAGAAGGATGGGCATGGGGAACGTTAAGCAAAGATCCAG 270
 60  G  H  Y  T  M  E  L  S  K  R  T  L  Y  V  T  K  H  E  E  G  W  A  W  G  T  L  S  K  D  P   89

271  GTCCACAGGGACATATGATCTTCATTCCTCTGAAATCTGGAAATTACATGATCAGTACACAAAAGTGGCCAGATTGGTTTTTATTCATGG 360
 90  G  P  Q  G  H  M  I  F  I  P  L  K  S  G  N  Y  M  I  S  T  Q  K  W  P  D  W  F  L  F  M  119

361  AATCAAGCAGTCATGGCTATATCCGGTGTTGGAAAGGAAATGTTGGGCCACAAGGAGTTTGGCACATTCTTTAAATAAGGATAAATCAAG 450
120  E  S  S  S  H  G  Y  I  R  C  W  K  G  N  V  G  P  Q  G  V  W  H  I  L  *                 143

451  CATTTTAATGTTGTTTCTTTATATAATGTACTGCTTTTATTTCCAGAAATATTGCAAGTTAATACGTCGGCATACTGCCTCAAACTGTTA 540
541  TTATTCAGGGCCGTAGCAGCCTAGGGGCCCGGGGGCCACAGCCCCCCAAATTTTTGGTGATTATAGGTGACTTGTAGAAAGCTGCAAGAA 630
631  ATAACCTTTTTAACCTTACTGAGGGTGCTTTTTTAGCCAATCTAGGAGGTCAAAATGTCAAAAAAATTTCCTTGCTCATCGAACCCAACC 720
721  GTGGTGGAGCCTCCACCGATAGTAATGTTCAAAATTTCTCAAATTCAGCCCCCCTCCCTCAAAATTGAGGCCCTTGCTACAGCACTTTTA 810
811  TTACCTTGCTAATGAAGCAGTGACTTATTTGTCTTAACAATTAACAAGGTTCTACGAATATATGTATGTCTTTTGTGCTAATTTTATAAA 900
901  GAGATTTTCATCCTAAAAAAA                                                                      922
```

Figure 6. 2 Nucleotide and deduced amino acid sequences of the MkTX-B*f*

The three cysteines are highlighted in blue boxes; the stop codon identified with an asterisk. Two possible polyadenylation signal sites and a polyA tail are underlined in black.

6.4.3 Primary structure comparison of M. kingi *cytotoxins and homologs from cubozoans.*

Searches using the BLAST programs at NCBI (Altschul *et al.* 1990; Gish & States 1993) revealed that MkTX-A and truncated MkTX-B*f* are representatives of the box jellyfish toxin family, which includes *C. fleckeri* CfTX-1 and -2 (43 and 45 kDa), *C. quadrigatus* CqTX-A (44 kDa), *C. rastoni* CrTXs (43 and 46 kDa) and *C. alata* CaTX-A (43 and 45 kDa). All of these proteins have been isolated, the cDNA and amino acid sequences elucidated, and the native biological activity experimentally

established (Brinkman & Burnell 2009). The proteins display potent haemolytic activity, are lethal to crayfish and mice (when tested), and possibly the cause of cutaneous inflammation in humans (Brinkman & Burnell 2007; Brinkman & Burnell 2008; Nagai 2003).

MkTX-A contains a signal peptide indicating that following synthesis, the protein is transported to a specialized compartment, in this case, the nematocysts. Sequence comparisons indicate that the truncated MkTX-B*f* cDNA clone lacks the nucleotide sequence required to encode ~ 296 N-terminal amino acids, however, by similarity it is also predicted to be a secreted product.

The alignment in Figure 6.4 compares the mature polypeptide of MkTX-A and MkTX-B*f* with the five published cytolysins from cubozoans. Apparent protein divergence was found among the group and it forced the introduction of several indels in the alignment. The alignment takes account of the chemical type likeness and PSI-BLAST pairwise similarity e values (Altschul *et al.* 1990). Figure 6.4 indicates a general trend exposing five considerations:

i) There are discrete blocks of conserved or similar type amino acids present across all members
ii) There is greater similarity between the N-terminal ends compared to the C-terminal ends of the proteins
iii) The number and distribution of cysteine residues varies between chirodropid and carybdeid cytolysins
iv) MkTX-A is the most divergent sequence and relates more closely to the chirodropid cytolysins than the carybdeid cytolysins
v) The MkTX-B C-terminus fragment is clearly more closely related to CrTXs and CaTX-A than to CqTX-A and the CfTXs. The sequence alignment shows that MkTX-A and MkTX-B differ the most between each other.

Figure 6. 3 ClustalW alignment of seven homologous cytotoxic proteins.

Light grey shading indicates chemically similar amino acids and dark grey highlights identical residues. Bold letters in the consensus line indicate residues that are highly conserved in all members of the box jellyfish toxin family. Cysteine residues are boxed in red and the red line links two of the cysteine residues that are predicted to form a disulphide bond (DISULFIND, Ceroni *et al.* 2006).

The presence of cysteine residues in the cubozoan cytolysins indicates that the products may form disulphide bonds. High confidence scores for cysteine bonding were retrieved (DISULFIND), for MkTX-A and -B and also for all of the other cubozoan toxins. The best connectivity pattern score in the case of MkTX-A is indicated in Figure 6.3. Predictions of disulphide bonding in cubozoan proteins correlate in part with the experimental results of Brinkman & Burnell (2007) where the apparent molecular masses of CfTX-1 and -2 extracted from *C. fleckeri* nematocysts remained unchanged in both reducing and non-reducing conditions, thus suggesting that disulphide bonds (if any) in the molecules would be intramolecular (or located at the molecule's surface).

In comparison to other cubozoan toxins, the cysteine residue in MkTX-A at position 211 is also present in CrTX within the conserved region {YCQ}. The cysteine at position 214 in MkTX-A is found in both CfTX-1 and -2 in the conserved sequence {LCSLR}. The cysteine at position 330 is also highly conserved, while the last cysteine (at position 370) is not. MkTX-A is also the only protein that possesses two closely associated cysteines flanked by K and R.

In anemones, components displaying haemolytic, lethal, and in some cases, cardiotoxic activity, have been detected in small cysteine-free peptide toxins, that are rich in basic and hydrophobic amino acids (Turk 1991). These include magnificalysins I and II from *Heteractis magnifica* (Khoo *et al.* 1993) cytolysin III (Lanio *et al.* 2001) and Sticholysins I and II (St I and St II) from *S. helianthus* (Alvarez *et al.* 2003). Site directed mutagenesis studies have shown that a positive charge is required for hemolytic activity and, at least in the case of equinatoxin, lysine 77 is required for binding to lipid membranes (Anderluh *et al.* 2000a). In the cubozoan cytolysins, basic amino acids are not only abundant but also highly conserved (Figure 6.5). Hence, it would be interesting to determinate whether positively charged residues play a role in the biological activity of cubozoan toxins.

6.4.4. Related proteins in Hydra magnipapillata

PSI-BLAST (Altschul *et al.* 1990) searches also identified five recently released sequences (February 2009) from *H. magnipapillata*, which are related with various degrees of similarity to members of the box jellyfish toxin family. The hydra proteins were named HmagTX1 (XP_002162997.1), HmagTXA (gi|221120315), HmagHyPro1 (gi|221113110), HmagHyPro2 (gi|221113112) and HmagHyPro3 (gi|221131661; Figure 6.5). Only HmagHyPro1 and HmagTX1 correspond to secreted protein precursors. In 2005 an EST analysis of *H. magnipapillata* reported certain tags to be similar to the box jellyfish cytolytic family (Sher *et al.* 2005b), but no protein characterization was conducted. It is likely that the authors were referring to those sequences mentioned earlier in the paragraph.

A

	MkCTX-A	CqTX-A_	CrTXs	CaTx-A	CfX-1	CfX-2	HmagTX-A	HmagTX-1	HmagHyPro1	HmagHyPro2	HmagHyPro3
Length (a.a)	434	443	432	440	436	445	473	440	661	804	578
Masses kDa	**48.55**	**49.69**	**47.34**	**49.08**	**49.14**	**49.88**	**53.14**	**49.61**	**72.71**	**88.59**	**65.91**
Isoelectric point	8.79	8.63	9.21	8.84	8.58	7.8	4.76	5.89	8.41	8.69	5.33
Aliphatic index	87.32	87.99	82.01	81.13	90.06	89.16	85.20	95.88	78.89	86.49	100.31

B

	MkTX-A	CfTX-1	CfTX-2	CqTX-A	CaTX-A	CrTX-A	HmagTX-A	HmagTX-1	HmagHyPro1	HmagHyPro2	HmagHyPro3
MkTX-A	100	**24.6**	**26.3**	**24**	17.4	19.4	**13.5**	**12.9**	10.8	9.7	7.1
CfTX-1	**46.7**	100	72.1	92.6	19.4	46.8	**12**	**23.8**	20.4	19.5	21.5
CfTX-2	**46.9**	82.9	100	83	21.6	45.2	**12.8**	**22.5**	21.1	19.3	21.1
CqTX-A	**45.8**	87.6	72	100	19.8	24	**12**	**13.2**	12.2	11.8	9
CaTX-A	36.6	39.1	40.9	38.6	100	60.3	**32.2**	**25.3**	21.6	19.1	22.1
CrTXA	35	24.9	23.8	45.7	42.6	100	**15.9**	**12.5**	10.4	9.9	8
HmagTX-A	30.8	27.9	29.8	27.9	15.7	33.4	100	**23.1**	20.3	17.8	19.3
HmagTX-1	24.4	14	12.8	23.4	12.6	25.1	**10.8**	100	25	33.3	7.5
HmagHyPro1	19.9	11.8	11.8	20.8	11.3	21.1	9.2	32.3	100	72.6	18.5
HmagHyPro2	17.5	11.3	11.1	20	10.4	19.4	7.7	25	68.4	100	7.9
HmagHyPro3	20.2	8.8	9.6	21	8.7	20.6	5.9	18.8	8.1	16.9	100

C

	MkTX-B	MkTX-A	CfTX-1	CfTX-2	CqTX-A	CaTX-A	CrTXs	HmagTX-A	HmagTX-1
MkTX-B	100	**13.9**	16.4	17.4	16.4	**63.9**	**68.1**	17.3	8.5
MkTX-A	**31**	100	28.1	30.2	30.8	30.4	32.7	10.9	8
CfTX-1	35.2	11.9	100	40.1	82.6	31.9	42.5	8.2	10.8
CfTX-2	34.8	16.4	57.9	100	63	36	39.5	9.8	14.8
CqTX-A	35.2	14.5	74.5	45.9	100	31.2	41.9	10.3	13.2
CaTX-A	**50.7**	***14.6***	16.9	19.9	17.5	100	54.5	15.4	7.9
CrTXs	**54.2**	***14.5***	21.2	19.8	21.2	66.9	100	14.1	7.3
HmagTX-A	**30.2**	**20.6**	18.7	20.2	20.6	30.2	28.8	100	13.6
HmagTX-1	19.5	19.1	23.4	27.5	25.7	17.1	19.4	7.4	100

D

	MkTX-A		CfX-1		CfX-2		CqTX-A_		CrTXs		CaTx-A		HmagTX-A		HmagTX-1		HmagHyPro1		HmagHyPro2		HmagHyPro3	
	Frequency	%																				
Leu **(L)**	44	**10.1**	46	**10.6**	49	**11**	48	**10.8**	33	**7.6**	35	**8**	43	**9.1**	53	**12**	52	**7.9**	74	**9.2**	67	**11.6**
Lys **(K)**	40	**9.2**	44	**10.1**	41	**9.2**	42	**9.5**	26	**6**	28	**6.4**	28	**5.9**	44	**10**	63	**9.5**	80	**10.0**	41	**7.1**
Ser **(S)**	37	**8.5**	39	**8.9**	43	**9.7**	42	**9.5**	44	**10.2**	39	**8.9**	35	**7.4**	38	**8.6**	63	**9.5**	71	**8.8**	46	**8.0**
Gly **(G)**	32	**7.4**	26	**6**	28	**6.3**	29	**6.5**	36	**8.3**	36	**8.2**	31	**6.6**	28	**6.4**	60	**9.1**	63	**7.8**	34	**5.9**
Thr **(T)**	31	**7.1**	17	**3.9**	23	**5.2**	19	**4.3**	21	**4.9**	31	**7**	18	**3.8**	11	**2.5**	27	**4.1**	40	**5.0**	23	**4.0**
Val **(V)**	29	**6.7**	32	**7.3**	27	**6.1**	32	**7.2**	26	**6**	27	**6.1**	25	**5.3**	29	**6.6**	44	**6.7**	48	**6.0**	42	**7.3**
Glu **(E)**	28	**6.5**	26	**6**	28	**6.3**	27	**6.1**	20	**4.6**	19	**4.3**	22	**4.7**	30	**6.8**	43	**6.5**	48	**6.0**	53	**9.2**
Ile **(I)**	27	**6.2**	25	**5.7**	26	**5.8**	22	**5**	28	**6.5**	28	**6.4**	32	**6.8**	29	**6.6**	39	**5.9**	54	**6.7**	43	**7.4**
Pro **(P)**	22	**5.1**	11	**3.7**	10	**3.6**	16	**3.6**	16	**3.2**	14	**2.5**	13	**2.7**	13	**3**	25	**3.8**	28	**3.5**	12	**2.1**
Tyr **(Y)**	22	**5.1**	21	**4.8**	18	**4**	19	**4.3**	16	**3.7**	24	**5.5**	30	**6.3**	15	**3.4**	28	**4.2**	31	**3.9**	28	**4.8**
Asp **(D)**	20	**4.6**	29	**6.7**	28	**6.3**	28	**6.3**	21	**4.9**	26	**5.9**	42	**8.9**	30	**6.8**	37	**5.6**	44	**5.5**	30	**5.2**
Ala **(A)**	18	**4.1**	23	**5.3**	26	**5.8**	24	**5.4**	41	**9.5**	33	**7.5**	38	**8**	18	**4.1**	39	**5.9**	57	**7.1**	29	**5.0**
Asn **(N)**	18	**4.1**	19	**4.4**	15	**3.4**	17	**3.8**	23	**5.3**	18	**4.1**	33	**7**	25	**5.7**	37	**5.6**	37	**4.6**	31	**5.4**
Phe **(F)**	18	**4.1**	24	**5.5**	24	**5.4**	25	**5.6**	15	**3.5**	19	**4.3**	10	**2.1**	20	**4.5**	22	**3.3**	30	**3.7**	27	**4.7**
Gln **(Q)**	15	**3.5**	16	**3.7**	18	**4**	19	**4.3**	13	**3**	14	**3.2**	23	**4.9**	23	**5.2**	21	**3.2**	23	**2.9**	17	**2.9**
Arg **(R)**	14	**3.2**	14	**3.2**	16	**3.6**	18	**4.1**	23	**5.3**	23	**5.2**	21	**4.4**	14	**3.2**	21	**6.2**	21	**2.6**	28	**4.8**
His **(H)**	7	**1.6**	5	**1.1**	6	**1.3**	4	**0.9**	10	**2.3**	9	**2**	1	**0.2**	1	**0.2**	4	**0.6**	4	**0.5**	2	**0.3**
Met **(M)**	5	**1.2**	8	**1.8**	5	**1.1**	6	**1.4**	11	**2.5**	10	**2.3**	12	**2.5**	10	**2.3**	11	**1.7**	15	**1.9**	18	**3.1**
Cys **(C)**	4	**0.9**	5	**1.1**	5	**1.1**	5	**1.1**	4	**0.9**	4	**0.9**	9	**1.9**	7	**1.6**	16	**2.4**	23	**2.9**	4	**0.7**
Trp **(W)**	3	**0.7**	1	**0.2**	3	**0.7**	1	**0.2**	7	**1.6**	6	**1.4**	7	**1.5**	2	**0.5**	9	**1.4**	13	**1.6**	3	**0.5**

Figure 6. 4 Sequence statistics of evolutionary-related cnidarian cytotoxins

(A) Protein characteristics; (B) Identity and similarity matrix of mature proteins in comparison to MkTX-A; (C) Identity and similarity matrix of the C-terminus of the cubozoan cytotoxic proteins and related *H. magnipapillata* proteins. Matrix constructed by MacVector using slow mode pairwise alignment parameters (Open Gap Penalty = 10.0, Extend Gap Penalty = 0.1, Delay Divergent = 40, Gap Distance = 5) and using blossum as a similarity matrix (MacVector). (D) Amino acid frequency and percentage/composition in mature products. Negatively charged residues (Asp and Glu) are indicated in green and positively charged residues (Arg and Lys) are marked in purple. The abundance of the aromatic amino acid Tyr is highlighted in pink, and the use of cysteine residues is shown in orange.

Although HmagTXA and HmagTX1 are related to box jellyfish cytotoxins they differ from each other. For example, in contrast to HmagTXI, HmagTXA is unlikely to be confined in hydra nematocysts because it is synthesized as a non-secreted product. It is important to recognize that not all cnidarian toxins originate from nematocysts. There are reports of a pore-forming toxin of non-nematocyte origin in a hydra species (Zhang *et al.* 2003) as well as other cnidarian defence response genes (Matveev *et al.* 2007). Thus HmagTXA may be a non-nematocyst origin protein able to cause cytotoxic effects in prey.

6.4.5 Secondary structure predictions of mature MkTX-A and MkTX-B and comparison to the other cytotoxins from Cubozoa and putative toxins from **H. magnipapillata**

The secondary structure of MkTX-A and -B as well as other cubozoan cytolysins were examined using a number of different methods including the MacVector toolbox: Chou Fasman, Robson Garnier, PSIPRED V2.6 (Jones 1999). Specifically, MkTX-A was predicted to be a combination of α-helices and β-strands all linked by coils on average 5 to 6 residues and up to 15 residues in length. The first 320 amino acid chain is dominated by 15 predicted α-helices and, preceding the last of these, a small β-strand is predicted. In contrast, the C-terminal (last 114 residues) of MkTX-A has ten predicted β-strands linked by long coil regions; these large loops may promote the C-terminus plasticity. Two out of five cysteines are located in the C-terminal region. In contrast, according to primary sequence analyses (PSIPRED and MacVector), MkTX-B constitutes 10 β-strands. High values of confidence indicate a single α-helix located from residues 34 to 41. Cysteines in positions 31 and 44 of MkTX-B were suggested by one method (DISULFIND) to form a disulphide bridge. In all proteins the third and the fourth cysteine residues are

predicted to connect the α- helix bundles with the β-sheet domain. (Compare Figures 6.1 and 6.3 in which the cysteine residues are highlighted).

6.4.5.1 Amphiphilic α-helices and transmembrane topology

According to amphiphilic helices plots and membrane topology prediction algorithms *e.g.* PHD, (Rost 1996), Jpred (Cole *et al.* 2008), all cubozoan cytotoxins, HmagTXA and HmagTX-I have at least one small amphiphilic region predicted in an alpha-helix region. In the case of MkTX-A, there may be two small amphiphilic helices (see Figure 6.5), followed by three transmembrane regions; the first is long enough to span a biological membrane. Although the amphiphilic and/or transmembrane helices of proteins have only a small number of shared identities between homologs, it is clear that these regions exist and putatively have similar function. Previous researchers have suggested that the amphiphilic α-helix may be involved in the insertion of cubozoan cytotoxins in lipid membranes (Nagai 2003), however this hypothesis has not been experimentally proven. Alternatively, the amphiphilic area may have an anchoring function and as suggested in (Brinkman & Burnell 2007) an adjacent transmembrane-spanning region could potentially penetrate the membrane.

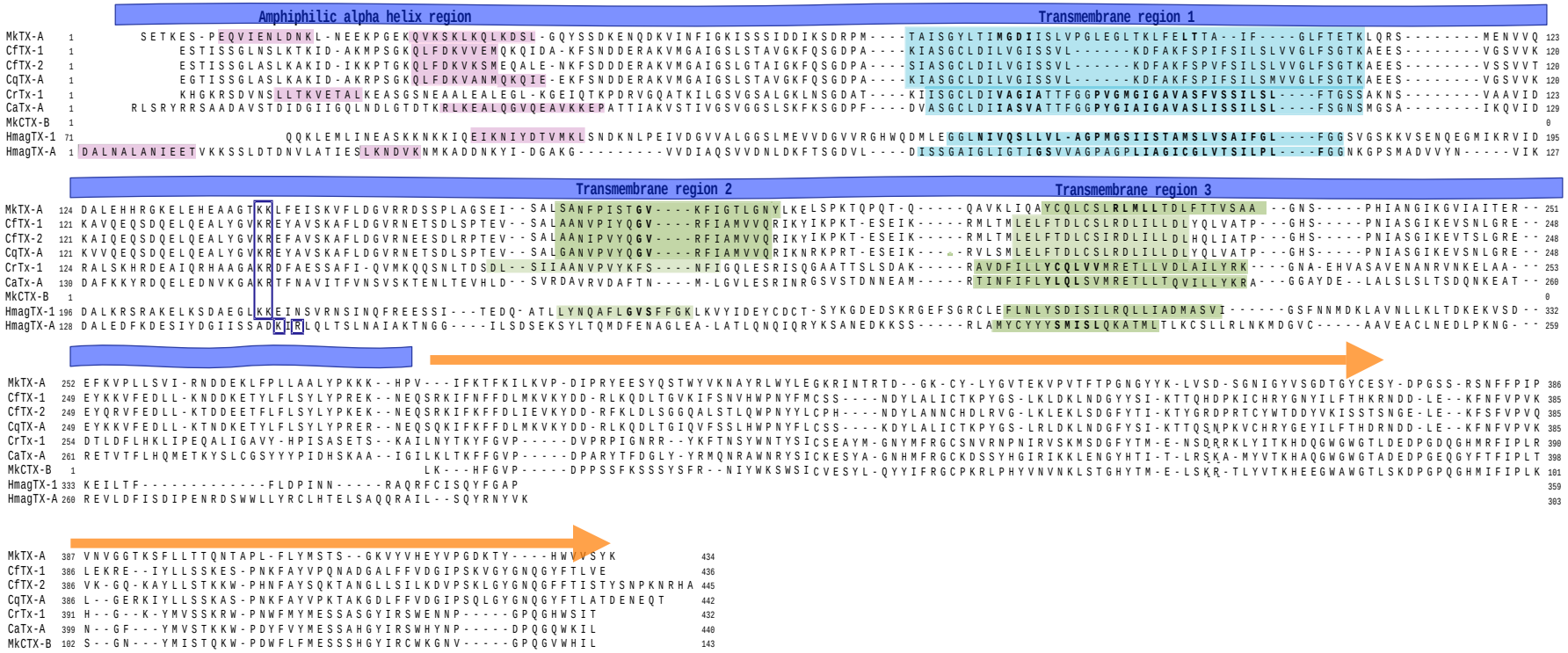


Figure 6.5 Prediction of transmembrane regions in seven cubozoan cytotoxins and two related proteins from *H. magnipapillata*

The blue and orange ribbons above the sequence alignments indicate the α-helix bundle and the C-terminal β-domain, respectively. Residues highlighted in pink correspond to amphiphilic α-helices. The alignment section in blue shows the most conserved and long transmembrane region. Two other small predicted transmembrane regions are highlighted in green. A conserved dyad of basic amino acids is boxed in purple.

6.4.5 General three-dimensional architecture of cubozoa and hydrozoan related proteins.

There are no three-dimensional structures available for the cubozoan proteins yet and the modelling templates used here were reached by searching the Protein Database (PDB) and the Structural Classification of Proteins database (SCOP) (Murzin *et al.* 1995). The structural and functional domains retrieved by protein structure data bases - Superfamily 1.69- (Gough 2001; Wilson 2009) are given in Table 6.1. This information provides hints on regions of the box jellyfish lysins with some degree of remote structural similarity to elucidated protein domains. Some of the predictions correspond to domains in known toxins, while other predicted domains have no apparent toxic properties.

3D models of the cubozoan cytotoxic proteins were predicted based on the chemical properties of amino acids substitutions from crystallized molecule templates SAM_TO8, I-TASSER (Katzman *et al.* 2008; Yang 2008). The general tertiary structure trends for the proteins are illustrated in Figure 6.7. A clear bias for two domains is apparent, as previously suggested by Brinkman & Burnell, 2009. The larger N-terminal domain is dominated by the presence of α-helical structures whereas the C-terminal domain is smaller and comprised mostly of β sheets.

Table 6.1 Structural and functional SCOP domains with remote structural homology to regions of cytotoxic jellyfish proteins.

Prot ID	Region	Superfamily	E value
MkTX-A	28-76	Prefoldin	6.41-03
MkTX-A	227-275	δ-Endotoxin –insecticidal-, N-terminal domain	2.42-01
MkTX-A	85-275	δ-Endotoxin –insecticidal-, N-terminal domain	3.22-01
MkTX-A	6-447	Tricorn protease N-terminal domain	3.23-01
CfTX-1	232-353	Cytokine (Interleukin-1 (IL-1)	1.10-05
CfTX-1	198-246	CheW-like	0.71
CfTX-2	21-57	Hypothetical protein AF0491, middle domain	0.029
CfTX-2	229-337	Cytokine (Interleukin-1 (IL-1)	2.4-05
CfTX-2	205-239	DNA breaking-rejoining enzymes	0.14
CfTX-2	92-154	δ-Endotoxin -insecticidal-, N-terminal domain	1
CqTX-A	232-344	Cytokine	4.30-08
CqTX-A	18-63	UDP-Glycosyltransferase/glycogen phosphorylase	0.13
CqTX-A	178-246	CheW-like	0.71
CrTXs	356-449	PK beta-barrel domain-like	0.58
CaTX-A	97-169	δ-Endotoxin -insecticidal-, N-terminal domain	0.0017
MkCTX-B	91-142	Actin-crosslinking proteins	8.82-02
HmagTXA	48-124	Mannose-binding lectins	6.13-02
HmagTXA	118-375	αβ-Hydrolases	1.48-01
HmagTX1	21-435	Spectrin repeat	4.26-02
HmagHyprot1	447-495	Kringle-like	7.01-06
HmagHyprot1	325-437	γ-Crystallin-like	6.90-03
HmagHyprot1	186-217	Rad51 N-terminal domain-like	8.75-02
HmagHyprot1	15-669	PLC-like phosphodiesterases (Glycerophosphoryl diester phosphodiesterase)	2.03-01
HmagHyprot2	452-565	γ-Crystallin-like (Crystallins/Ca-binding development proteins)	2.19-03
HmagHyprot2	575-622	Kringle-like (Fibronectin type II module)	4.37-03
HmagHyprot2	17-29	N-terminal Zn binding domain of HIV integrase	4.77-02
HmagHyprot2	316-778	Metalloproteases (“zincins”), catalytic domain	7.13-02
HmagHyprot2	22-748	γ-Crystallin-like	1.21-01
HmagHyprot3	290-370	δ-Endotoxin -insecticidal-, N-terminal domain	1.36-01
HmagHyprot3	248-301	QueA-like (Pfam 02547)	1.64-01
HmagHyprot3	28-125	Transferrin receptor ectodomain, apical domain	4.14-01

The 3D model of MkTX-A was constructed based on the structure of δ-endotoxin CRY3BB1 of *Bacillus thuringiensis* (PDB 1ji6A) with an e value of $5.58e^{-02}$. MkTX-A was predicted to have two structural domains, an N-terminal domain containing at least 12 α-helices and a β-sheet C-terminal domain composed of six β-strands. The two modules may be held together by a thiol bond between cysteines (residues C330 and C370) as seen in the structural model.

The best template for MkTX-B corresponded to the hemagglutinin component (HA1) from type C botulinum neurotoxin from *C. botulinum* (PDB 1qxmA). HA1 is also structurally similar to the lectin b-chain from the deadly plant toxin, ricin and belongs to the same "Ricin B-like family and superfamily of lectins" (SCOP b.42.2.1). This family displays a beta trefoil topology and all members are beta proteins. A similar structure is expected for the C-terminus of MkTX-B that is predicted to be composed of about 10 β-sheets and at least one α-helix. Similar structural predictions (I-TASSER) were obtained for the C-terminus of CaTX-A and CrTX-A (but not for CfTX-1, CfTX-2, CqTX-A or MkTX-A) and in both cases remote structural homology to Abrin A, a natural poison that is found in the seeds of plants and is similar to ricin (Tahirov *et al.* 1995), was detected. These results indicate that the C-terminal fragment of MkTX-B could be involved in binding to cell surface carbohydrates. According to consensus predictions of five functional homologs (1qxmA, 2ehiB, 2ehiB, 2ehmB, 2ehnB), the putative binding site residues of MkTX-B include:(Cys:31(2), val:32(2), glu:33(2), ser:34(2), arg:42(3), Pro:45(2) and Pro:49(2)) all of which are related to the HA1 hemagglutinin component (Chains A and B) of botulinum type C progenitor toxin.

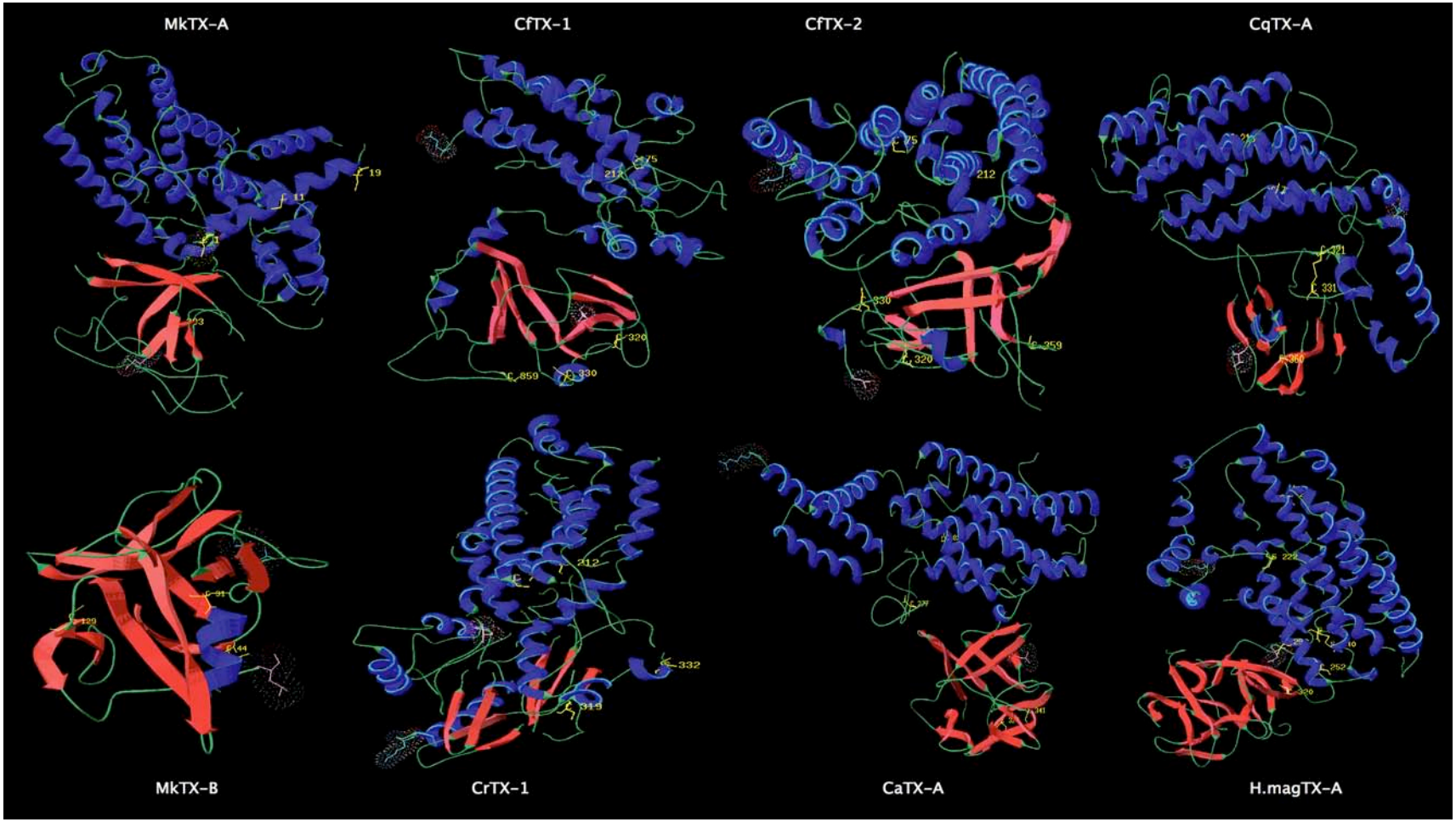


Figure 6. 6 Three dimensional modelling of cubozoan cytolysins and HmagTXA

Predictions were done using the SAM09 algorithm by residue chain substitution of the best scored template and modelling in Swiss-PDBviewer. In the case of MkTX-A the best alignments was 1ji6A-5.5766e-02 and the template has the SCOP Domains: b.18.1.3, b.77.2.1and f.1.3.1. for MkTX-B the PDB 1qxmA.

6.4.6 Evolutionary divergence of the Cubozoan cytolysin family

Comparative sequence alignments of mature MkTX-A using PSI-BLAST identified significant homology with other Cubozoa cytotoxins, HmagTXA and to a lesser extent to HmagTXI and hypothetical *Hydra* proteins. HmagTX-A showed sequence similarity with 16-18 kDa salivary gland proteins (secreted) from several species of mosquito and the gene product of GK25225 from *Drosophila willistoni.* A high degree of similarity was also found with two C-terminal internal repeats of a hypothetical and non-secreted 564 residue long protein from mosquito (AaeL_AAEL004188a). Similarity between Cnidaria and arthropod sequences retrieved high ranking of e-values 7e-04 to 5e-08 indicating that they may have evolved independently from an ancestral common gene. The Cubozoa cytolytic proteins are not related to any other cnidarian pore forming cytolytic proteins reported so far from Cnidaria *e.g.* Actinoporins (Macek *et al.* 1994), β-pore forming toxins (Sher *et al.* 2005a) or equinatoxin; although the latter is similar to Cubozoa lysins in the sense that an N-terminal amphiphilic region is involved in membrane penetration (Malovrh *et al.* 2003).

Polypeptide sequences (24) were used to attempt reconstruction of the evolutionary history of the Cubozoa cytotoxin family. Twelve of the sequences used in these analyses were Cnidaria-derived including seven from Cubozoa. Arthropod sequences were included because simple comparison among related cubozoans are unlikely to be informative in terms of understanding the origins and evolution of box jellyfish hemolysins. Two approaches were considered: a Neighbour joining (NJ) and a maximum likelihood protein inferences. The results were similar using these approaches.

The NJ method divided the sequences into two main clusters (see in Figure 6.7 A); the small cluster contains four hypothetical arthropod sequences. The major cluster is divided into two groups: one comprises only arthropod members and the other is specific for cnidarians and indicates that the hypothetical *H. magnipapillata* protein and HmagTX1 are distantly related to the Cubozoa cytolytic family. The most

important feature of this tree is that the cytotoxins from Cubozoa and HmagTX-A from *H. magnipapillata* cluster together. The Cubozoa cytotoxins split into two clades arbitrarily named one and two. Clade one has three of the more divergent sequences in the group and these correspond to chirodropid species. MkTX-A is included in this branch even though it belongs to Order Carybdea. Sequence comparisons are reflected in the tree and indicate that MkTX-B is more closely related to CaTX-A and CrTXs than to MkTX-A.

A more confident topology tree of the Cubozoa phylogeny was achieved using multiple ClustalW protein alignment as seed to run the Maximum likelihood algorithm (Molphy 2.3). The resulting tree is shown in figure 6.7 B. In this analysis, HmagHyp3 was used as the baseline. The tree illustrates two clusters dividing completely the arthropod and the cnidarian sequences indicating that these proteins have evolved separately. According to Figure 6.7 B, the cnidarian cluster, grouped the hypothetical Hmag proteins and HmagTX. While in the other branch HmagTXA and the toxins from Cubozoa formed a well-supported group in which the cubozoan cytotoxins divide into two separate clades. The protein divergence between the clades may account for differential functional properties. However a more extensive cubozoan dataset may clarify the evolution of these novel and intriguing cnidarian toxic proteins.

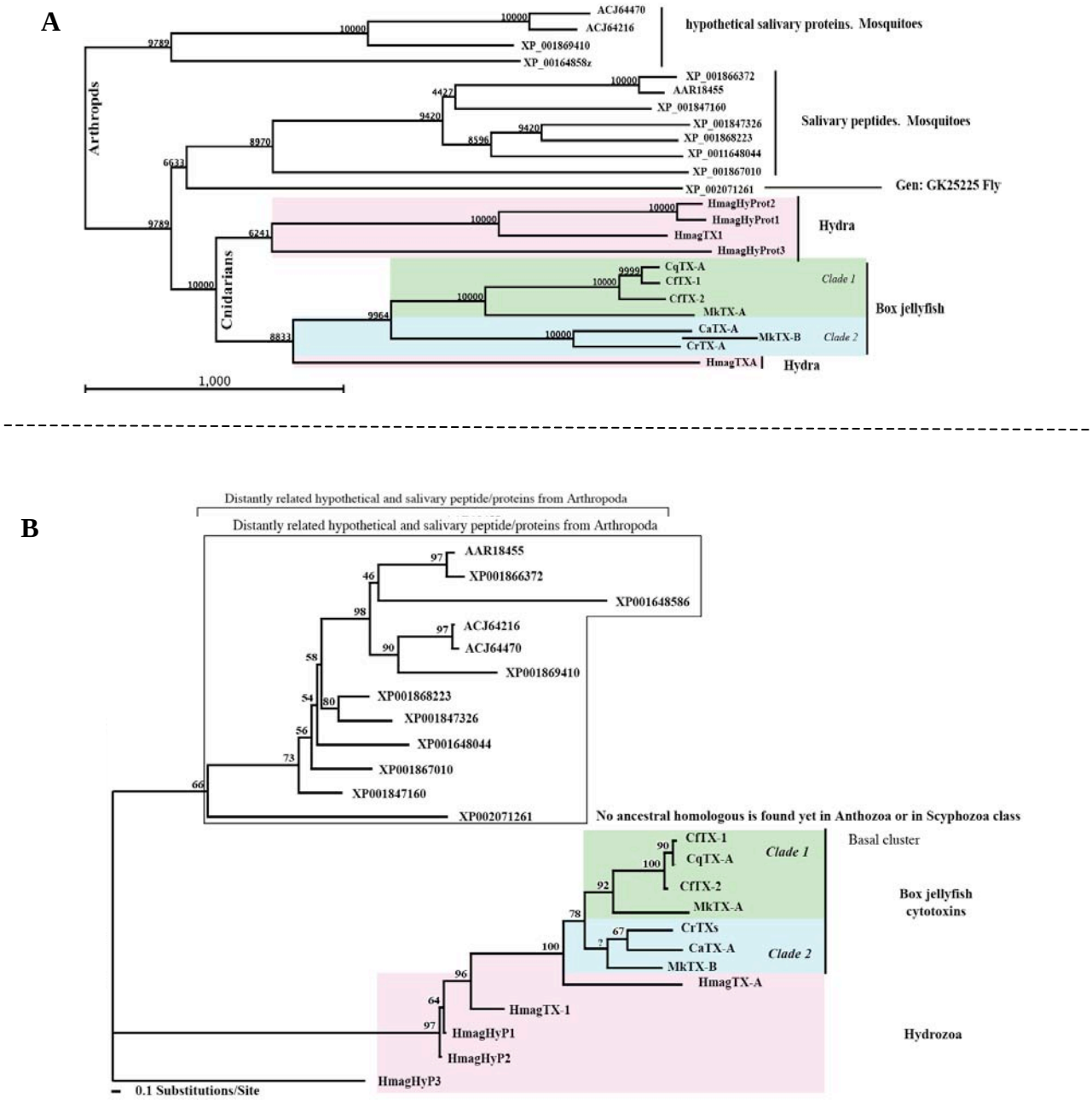


Figure 6.7 Two approaches to infer the phylogeny of Cubozoa toxins along with related *Hydra* and distantly related arthropod proteins.

A. Phylogenetic tree constructed with the following parameters: Open gap penalty of 10.0, Extend Gap Penalty of 0.1, Delay Divergent of 40, Gap Distance of 5 and using blossum as similarity Matrix. B. Maximum likelihood inference of protein phylogeny constructed using standard neighbour joining tree topology parameters. The bar indicates distances between sequences. Cnidaria sequences used in these analyses: CaTX-A Q9GNN8, CrTXs Q9GV72, MkTX-BFJ705829, CqTX-A P58762, CfTX-1 ABS30940, CfTX-2 ABS30941, MkTX-A FJ705827 and FJ705828, HmagTX1 XP_002162997, HmagTXA XP_0021616.

6.5 CONCLUSION

Through the course of this research, and in spite of numerous collecting exercises, it was not possible to collect sufficient numbers of *M. kingi*, to allow duplication of a simple venom profile analysis. In addition, the amount of nematocyst proteins fractionated by SDS-PAGE was insufficient to allow determination of the N-terminal or internal amino acid sequences which could be used to design degenerative primers. Two of the three available jellyfish samples were used to construct a cDNA library that was screened with antibodies and later randomly sequenced. The latter approach resulted in the identification of a partial cDNA clone encoding MkTX-A. Using a DNA hydridization assay, cDNA clones encoding mature MkTX-A and the C-terminal fragment of MkTX-B were obtained.

Discovery of MkTX-A and -B in a small data set provides evidence that cytolytic proteins are highly expressed in the nematocysts of Irukandji box jellyfish. This finding correlates to SDS-PAGE analyses of *C. barnesi* and *C. fleckeri* nematocyst proteins in which homologous 43 and 45 kDa cytolysins are abundant. Phylogenetic analyses indicate that MkTX-A and -B belong to the cytotoxic box jellyfish protein family. Protein alignments indicate that there are two protein groups within the family of seven members from *M. kingi, C. rastoni, C. alata, C. quadrigatus and C. fleckeri*. In addition, five related molecules from *H. magnipapillata* have been recently identified by comparative analyses of primary sequences. Phylogenetic comparison congregates one of these proteins (HmagTXA) into one of the Cubozoa clusters. The hypothetical *H. magnipapillata* protein, HmagTX-1 and the arthropod sequences are distantly related to the cubozoans and may have originated from a common ancestor gene and then evolved independently.

Similar 3D structures were predicted among the cubozoan cytotoxins and HmagTXA. The structures are characterized by the presence of an N-terminal domain dominated by α-helixes while the C-terminal is relatively small and composed of at least 10 β-strand regions. In addition, all the C-terminal domains of the box jellyfish toxins and HmagTXA are predicted to have at least one disulphide bridge indicating not only

protein flexibility but fold stability. The C-terminal domains of MkTX-B, CrTX-A and CaTX-A have significant structural similarity with the cytotoxins predicted to belong to the ricin β-like protein family superfamily of lectins (SCOP b.42.2.1).

The cDNA sequence data discussed in this chapter were mostly characterized on the basis of primary structure using bioinformatic tools available as stand-alone or online-based algorithms programs. Although, the results retrieved and analysed are of great value they are still predictions and should not be taken as factual. Since biological specimens of *M. kingi* have proved difficult to collect it is reasonable to suggest that further work on the identification and characterization of potentially toxic proteins of this jellyfish may have to rely on the expression of recombinant proteins. Appropriate vector system technologies will allow recombinant protein production studies of cytolysins from cubozoans

CHAPTER 7

Overall research panorama, achievements and expectations

7.1 RESEARCH OVERVIEW

The prospect of Irukandji box jellyfish, such an extremely rare, seasonal and tiny creature having such potent venom is both alluring and feared. Paradoxically, for years, sufficient amounts of material could not be obtained for detailed biochemical and functional characterization and for this reason this research adopted a genomic strategy in order to overcome sample supply, accelerate the discovery of genes encoding defence proteins and to understand the general biology at the molecular level of two medically important cubozoan jellyfish: *C. barnesi* and *M. kingi.*

Irukandji jellyfish are well known to cause a complex and variable syndrome in humans. Hundreds of Irukandji syndrome cases are reported each year and two fatalities have occurred in North Queensland, one of them attributed to *M. kingi* (Gershwin 2005b). The first Irukandji jellyfish (*C. barnesi*) was discovered in Australian waters more that 56 years ago (Flecker 1952) and we now know that several species of box jellyfish in Australia and around the world can cause Irukandji syndrome in humans (De Pender *et al.* 2006; Gershwin 2006a; Gershwin 2005; Grady & Burnett 2003; Kinsey *et al.* 1988; Little *et al.* 2006a; Pommier *et al.* 2005). Unfortunately there is no current antivenom for Irukandji jellyfish and *C. fleckeri* antivenom has not been of use in treating Irukandji stings.

Most of the research results reported in this thesis are novel and were obtained with a restricted amount of biological material, emphasizing the importance of gathering molecular and putative functional information of defence mechanism genes that could allow further work for our research group and others.

7.2 FINAL REMARKS AND CONCLUSIONS.

7.2.1 C. barnesi *and* M. kingi *complementary cDNA libraries*

Two cDNA expression libraries were constructed and results collected during the course of this research demonstrated that both libraries were well represented, non redundant and with average insert size. In addition, it is important to point out that the *C. barnesi* library contained chimeric clones and this was due to the short cut used to construct the library given the paucity of starting material. Nevertheless, these clones are not over- represented.

7.2.2 Immunological screening of expression cDNA libraries

As part of an initial strategy an array of antibody probes was investigated for antigenic reaction towards box jellyfish nematocyst proteins from *C. barnesi* and *C. fleckeri*. Although different levels of reactivity were clearly observed, genuine antigenic reactions were observed towards species-specific polyclonal antibodies produced in mice, human sera, and CSL *C. fleckeri* antivenom (see figure 2.1 strip 7 and 10 as well as the dot blot in figure 2.2). It seemed then particularly attractive to exploit this feature by screening the Irukandji cDNA expression libraries for antigenic recombinant proteins. This approach was reinforced by the fact that several research groups have successfully implemented this strategy using either specific/unspecific polyclonal or monoclonal antibodies, resulting in identification of cDNA sequences encoding toxic proteins from various sources including plants (Mittermann *et al.* 2005), fungi (Saxena *et al.* 2003), parasitic invertebrates (Bachrach *et al.* 1997), honeybees, hornets (Fang *et al.* 1988), wasps (Jones *et al.* 1992) and snakes (Welton & Burnell 2005).

In this research however, no antigenic clones were isolated from the Irukandji cDNA libraries screened with human sera or species specific polyclonal antibodies raised in mice and the vast majority of clones that were antigenic to CSL box jellyfish antivenom were divided into house-keeping, cell-cell communication, transcription factors and miscellaneous genes. The resulting data clearly identified unwanted

background clones and did not detect clones of genes encoding toxins similar to other metazoans or novel genes. Therefore, it was concluded that complex antibodies are not suitable to detect Irukandji toxic components expressed in cDNA libraries. In spite of this initial outcome molecular data was acquired which was of value.

7.2.3 Expressed Sequence Tags

A random expressed sequence tag (EST) project using *M. kingi* cDNA library aimed to investigate the defence mechanism genes present in cubozoans. Surprisingly, in spite of the EST collection size, several defence mechanism genes were identified, those are highly expressed transcripts in adult Irukandji jellyfish, as only a small proportion of the transcriptome would have been sequenced. The most relevant tags discussed in the manuscript include

1.- ESTs 269 which share high homology with the two most abundant secreted lethal haemolytic proteins found in the nematocysts of other box jellyfish such as *Chironex fleckeri* - CfTX-1 43 and CfTX-2 45 kDa - (Brinkman & Burnell 2007), *Caribdea rastoni* -CrTXs 43–45 kDa-, *Caribdea alata* -CaTX-A 43 kDa- and *Chiropsalmus quadrigatus* -CqTX-A 44 kDa- (Nagai 2003).

2.- The occurrence in the collection of *MK-332,* a putative protein active in potassium channels.

3.- EST_906, which belongs to the stress response proteins exhibited homology with a novel dermatopontin-like toxin from the fire coral, *Acropora* sp (Iguchi *et al.* 2008).

4.- EST 45 and 310 were found to encode a secreted pathogenesis-related protein (PR) with a cysteine rich domain similar to the ShK toxic peptide. PR proteins are mostly (but not entirely) limited to plants and members have been reported in diverse phylogenetically unrelated species from bacteria, virus and animals (Geer *et al.* 2002).

5.- EST_63, a peptidase likely to be involved in both prey paralyses and digestion.

Remarkably, collateral relevant outcomes were achieved from the gene discovery project. Comparative gene analyses of the cuboza EST collection revealed that box jellyfish more closely resemble anthozoans than to hydrozoans, challenging long held views of cnidarians evolution. Importantly, it was observed that the cubozoa genetic pool is more similar to vertebrates (*e.g* mammalians) than to invertebrates like flies, nematodes or comb jellies (ctenophores); this is very interesting and has been previously reported for other cnidarian species (Kortschak *et al.* 2003). During the analysis of the cubozoans sequence data a small percentage of genes displayed high degree of similarities with non-metazoan species (e.g. plants, prokaryotes, viruses and even unicellular eukaryotes) were encountered. Non-metazoan-like genes detected in this study were ruled out to be from a contaminating source. Moreover, similar results from more complete data sets have been reported in cnidarians (Technau 2005) and this strongly supports the data presented in this thesis. Therefore it may be concluded that the EST cloning strategy was of great value in the discovery of cubozoan defence genes and at the same time provided a starting point to inspire advances in the research of cubozoan biology and evolution.

7.2.4 Cubozoan peptidases and peptidase inhibitors

Although it has never been reported for the class Cubozoa, it is well established that cnidarians host an array of proteases and protease inhibitors. In fact both protein groups are extensively studied and their occurrence has been reported in representatives of all kingdoms. These proteins play pivotal roles in a variety of biological processes including not only non-specific degradation of dietary proteins but are also involved in cell signalling, dictating cell behaviour during ontogeny, remodelling extracellular matrix, fertilization, immunity, apoptosis, protein processing, fibrinolysis and blood coagulation (David *et al.* 2005; Puente *et al.* 2003 ; Rawlings *et al.* 2008; Rojas & Doolittle 2002; Seipp *et al.* 2006; Technau *et al.* 2003). Chapter five was devoted to analyse a small fraction of what would be the degradome and the complement of protease inhibitor genes in *M. kingi.* The EST collection included members of four of the five existing families of peptidase: serine and metallo (S1A, M12 being the most abundant peptidases), followed by cysteine, threonine peptidases; but no aspartate or glutamate peptidases were identified although

members of this families occur in *N. vectencis*. Searches in the MEROPS data base indicated that the starlet anemone may lack only the small peptidase glutamate family and this may also be expected for box jellyfish. On the other hand, nine genes encoding peptidase inhibitor members were found. Our study demonstrated that adult Irukandji box jellyfish abundantly express proteolytic-related transcripts. Three peptidases of interest were found and these were classified as chymotrypsin-like serine peptidases (S1family, Clan A) and named Cb-TSP-1, Mk TSP-1 and MK-ShK-TSP-2. MK-ShK-TSP-2 has a domain containing six cysteine residues with a high degree of similarity to the ShK toxin domain within the protective peptide; similar protein architecture is found in the chymotrypsin-like serine protease in a schyphozoan species reported by Rojas & Doolittle (2002). The toxic domain of these types of proteinases, at least in cnidarian members, could serve as a toxin to keep prey paralysed during digestion although this suggestion needs to be investigated experimentally.

Whole mount mRNA *in situ* hybridisation experiments were conducted to elucidate the gene expression patterns of two different trypsin-like serine proteases. The results indicated that Cb-TSP-1 and Mk TSP-1 genes are expressed in the gastrodermis and also in the cnidoblast cells and MK-ShK-TSP-2 is only expressed in cnidoblast cells. This finding indicates that this protein is recruited into the nematocsyt after it is secreted. The presence of these genes in the armament of box jellyfish is feasible and is supported by previous reports of chymotrypsin serine protease activity in the venom of the medusa, *R. nomadica* (Gusmani *et al.* 1997)

The protein architecture and phylogenetic comparisons among three representative Cnidarian proteases: NvTSP-ShK, Mk-ShK-TSP-2, *Aurelia aurita* serine proteases and other TSP-homologs in other metazoans species, revealed that although the serine protease domain is highly conserved across all animals no trypsin serine protease-containing ShK domain is found in other metazoan, fungi or virus. Therefore it is reasonable to conclude that the architecture observed in Mk-ShK-TSP-2 represents a Cnidarian invention.

SDS-PAGE and western blot analyses demonstrated that Cb-TSP-1 and Mk TSP-1 could not be expressed in bacteria. In contrast, expression of MK-ShK-TSP-2 (including the protective peptide) was achieved and the recombinant proteins were immunologically reactive to CSL box jellyfish antivenom.

7.2.5 Putative neurotoxins from C. barnesi *and* M. kingi

Neurotoxic effects in Irukandji box jellyfish stung humans is evident and have been well reported and postulated to be mainly the result of catecholamine storm release in the blood stream (Burnett *et al.* 1998; Tibballs 2006a). Consistently, cardiovascular effects of Irukandji (*C. barnesi* and *Alatina mordens*) venom extracts in animal models (*in vivo) and in vitro* studies have shown effects similar to those observed in humans indicating that a neural (Na^+) channel activator may be the cause of the massive release of transmitters (Ramasamy *et al.* 2005b; Winkel *et al.* 2005; Winter *et al.* 2008), however no gene, or the protein encoded, has been reported. In the search for neurotoxic encoding transcripts in the *M kingi* and *C. barnesi* cDNA libraries, primers were designed in accordance with the most common codon usage of Cubozoans (See chapter 2, Table 2.1) on the basis of an internal amino acid sequence previously characterized as a putative neurotoxin -Ken Winkel personal communication- (refer to chapter 4). Two populations of cDNAs were found and the proteins encoded are synthesized as a preproprotein and a proprotein named CbTX-1 and 2, respectively. Homologs to CbTXs were identified in the *M. kingi* EST collection and those are Mk-332-1 and -2. The precursor organization is characterized by peptidase recognition sites and along prepropetide regions ending in polar and charged residues, similar organization to those short chain K^+ neurotoxins, Na^+ neurotoxins (Moran *et al.* 2008b), pore forming toxins from anemones, nematocyst collagens (Anderluh *et al.* 2000b) and “Aurelin” (Ovchinnikova *et al.* 2006). CbTX and Mk-332 have a tandem arrangement of three to four domains containing in the first three domains the signature XCXD and three out of the six cysteine residues in the last nine amino acids with the pattern CXXXTCXXC. These domains overall resemble strongly the *S. heliantus* K^+ toxins and it is concluded that CbTX and Mk-332 belong to the ShK toxin family.

According to cDNA sequences, the transcripts seem to originate from a multicopy gene family that arose by duplications and domain shuffling events from an actinaria ShK-like minigene (Gendeh *et al.* 1997a) although genomic information needs to be secured in order to elucidate the genomic organization of CbTX and Mk-332 genes.

Differential gene expression by RNA *in situ* hybridisation was observed in adults versus juveniles of Irukandji box jellyfish. The spatial distribution of CbTX and Mk-332 were similarly expressed in specific regions specialized for digestion (lips, manubrio and gastric cavity) and prey capture (battery of nematocytes in tentacles) in juveniles and in adults only in the battery of nematocytes.

Preliminary experiments which involved injecting CbTXs recombinant proteins into cockroaches showed that CbTXs are able to induce neurotoxic effects in the form of long term and lethal paralysis. Moreover according to NTXProtein (Saha & Raghava 2007), it is predicted that CbTXs could play a role in producing paralysis by blocking neuromuscular transmission at the post synaptic site. According to the domain composition and the similarity with the K^+ channel toxins from anemones it is likely that CbTXs may modulate K^+ channel receptors. It is anticipated that Mk-332 isoforms could accomplish a similar function, although further research on the bioactivity of recombinant proteins should be carried further.

7.2.6 Irukandji box jellyfish lysins

The *M. kingi* EST collection facilitated the identification, isolation and molecular characterization of two closely related and highly expressed genes: MkTX-A and the MkTX-B*f* (only the C terminus). These proteins belong to the lethal, hemolytic pore forming cytolytic family from cubozoa (Brinkman & Burnell 2009), lack relationship with other non Cubozoa-cnidarian pore forming cytolysins reported so far but have significant similarity with five uncharacterized secreted or non secreted *H. magnipapillata* proteins and other salivary proteins from arthropods.

Primary structure comparisons showed that protein characteristics such as calculated masses, isoelectric points, aliphatic index as well as a bias for conservancy of basic

amino acid composition are shared among all box jellyfish lysins including MkTX-A and the MkTX-B*f*. In contrast to other pore forming toxins from cnidarians, the cubozoa lysins are predicted to be cysteine cross-linked polypeptides. The number and distribution of cysteine residues varies between cytolysins from chirodropid versus caribdeids species. Tertiary structure and three dimensional modelling predicted that MkTXA has a characteristic protein scaffold consisting of two structural domains disulphide bonded by the third and the fourth cysteine residues. The N-terminal contains a bundle of at least 12 α-helices whereas the C terminal domain is composed of 150 amino acid residues adopting a six β-sheet arrangement. Amphiphilic helices plots and membrane topology prediction algorithms for MkTXA indicate that the N-terminal α-helices bundle domain contains two small amphiphilic helices followed by three transmembrane regions; the first may be long enough to span a biological membrane. Similar topology has been previously identified in cubozoan toxins (Brinkman & Burnell 2009). The presence of two functional domains bonded by the third and the fourth cysteine residues may be also be common for all of the family members.

The amino acid sequence for MkTX-B is incomplete and is predicted to be missing approximately 296 residues from the N terminus. The MkTX-B*f* corresponds to the C terminal and was found to have structural similarity with ricin (a toxic lectin from plants); similar results were retrieved for CaTX-A and CrTX-A but nor for any of the chirodropid toxins nor for MkTX-A.

Phylogenetic inferences showed arthropod salivary proteins to be distantly related to cnidarians proteins that have been confirmed to be toxins. Box jellyfish cytolysins are more related to HmagTXA and share less homology to HmagTXI and the three *H. magnipapillata* hypothetical proteins found in GenBank. Maximum likelihood algorithms applied to the data retrieved sequence divergence among the seven homologous cytotoxic proteins from Cubozoa and separated them into two clades. Clade one comprised the most divergent protein: MkTX-A from the carybdeid *M. kingi* and the three of the chirodropid toxins: CfTX-1/2, CqTX-A from C. *quadrigatus* and *C. flecker,i* respectively. Clade two was exclusively formed from carybdeid proteins: CaTX-A, CrTX-A and MkTX-B (*C. alata*, *C. rastoni* and *M. kingi*). The

amino acid divergence observed in cubozoan cytotoxins is likely to reflect functional significance.

Finally, it should be stressed that SDS-PAGE analyses of *C. barnesi* and *M. kingi* nematocyst proteins identified the major venom components as a doublet of apparent molecular weight ranging from 43 to 45 kDa. Western blot analyses of these proteins revealed this doublet to be immunodominant towards species-specific polyclonal antibodies raised in mice against *C. barnesi* and *M. kingi* jellyfish nematocyst extracts. In addition the *C. barnesi* doublet of 43 and 45 kDa was also antigenic towards rabbit antibodies species-specific for *C. fleckeri* 43 and 45 kDa and CSL *C. fleckeri* antivenom. Collectively, these data demonstrated that the 43 and 45 kDa proteins observed in *C. barnesi* and *M. kingi* jellyfish nematocyst extracts are homologs to those toxins reported by Nagai (2003) and Brinkman & Burnell (2007).

A shortage of jellyfish samples drove the direction of the research reported in this thesis and was the stimulus to generate an EST library from which Mk-TX-A and B were isolated. In spite of multiple attempts it has not been possible to isolate the *C. barnesi* homologs from the cDNA library. Protein comparison matrices indicated that the percentage of similarity among cubozoan cytolysins is low (in the range of 10.2-22.7%) and this may explain why attempts to isolate full length clones failed.

7.3 FURTHER DIRECTIONS

If further research involving box jellyfish antivenom is undertaken it should be conducted with caution due to the low number of specific nematocyst protein antibodies present in the antivenom that may not indicate the complete set of the venom proteins present in box jellyfish. This assumption is supported by western blot analyses of native venom proteins which demonstrated that only a few nematocyst proteins were antigenic to the antivenom. The reason for this may be various including for example the production of antibodies with impure venom preparations *e.g* containing perhaps mucus, mesoglea and degraded proteins that may act as immunodominant agents, with low antigenicity index of some venom components.

Compared with the other three Cnidaria classes, the molecular characterization of cubozoan genes has been greatly neglected. We have generated good quality expression libraries and a small, but to date unique, EST collection for an Irukandji box jellyfish. Note that the results presented in this thesis have emphasized the presence of defence mechanism genes however due to space and time limitations some other genes related to stress, immune response along with relevant developmental genes were not discussed in detail but nevertheless were present in the transcript tags collection. The complete sequence data obtained during this research will be loaded onto the public data base in the near future and this may allow further studies by the cnidarian-based research community.

Despite the research reported in this thesis and previous pharmacological studies, our knowledge of biologically-active components in Irukandji venom is still incomplete. Additional researchers will face the same difficulties encountered during this research *i.e.* difficulties in obtaining sufficient and pure venom proteins. It is my hope that the research reported in this thesis may stimulate interest in the benefits that genomic approaches can offer. I am confident that with the limited availability of jellyfish samples, molecular cloning strategies are practical and cost effective and without doubt will strengthen the current knowledge of cubozoan biology and, in time, help to unravel the properties of toxic proteins that will assist in the development of new treatments for envenomation as well as be useful in species identification.

REFERENCES

ABC News. 2006 Jellyfish Sting Kills 7-Year-Old: Australian Broadcasting Corporation. www.abc.net.au/news/stories/2006/01/08/1543650.htm.

ABC News. 2007 Boy's Death Prompts Stinger Warning.: Australian Broadcasting Corporation. www.abc.net.au/news/stories/2007/11/13/2089350.htm.

Abe, Y., Chinzei, T., Isoyama, T., Ono, T., Mochizuki, S., Saito, I., Guba, P., Karita, T., Sun, Y. P., Kouno, A., Suzuki, T., Baba, K., Mabuchi, K. & Imachi, K. 1999 Present status of the total artificial heart at the University of Tokyo. *Artificial Organs* **23**, 221-228.

Achatz, G., Oberkofler, H., E., L., Birgit, S., Unger, A., Kandler, R., Ebner, C., Prillinger, H. & Kraft, D. 1995 Molecular cloning of major and minor allergens of Alternaria alternata and *Cladosporium herbarum*. *Molecular Immunology* **32**, 213-227.

Adachi, J. & Hasegawa, M. 1996 MOLPHY version 2.3: program for molecular phylogenetics based on Maximum likelihood. *Computer Science Monographs* **281**, 1-150.

Adams, M. D., Kelley, J. M. & Gocayne, J. D. 1991 Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–6.

Adhikari, D., Samanta, S. K., Dutta, A., Roy, A., Vedasiromoni, J. R. & Sen, T. 2007 In vitro hemolysis and lipid peroxidation-inducing activity of the tentacle extract of the sea anemone (Paracondylactis indicus Dave) in rat erythrocytes. *Indian Journal of Pharmacology* **39**, 155-159.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Alvarez, C., Casallanovo, F., Shida, C. S., Nogueira, L. V., Martinez, D., Tejuca, M., Pazos, I. F., Lanio, M. E., Menestrina, G., Lissi, E. & Schreier, S. 2003 Binding of sea anemone pore-forming toxins sticholysins I and II to interfaces Modulation of conformation and activity, and lipid-protein interaction. *Chemistry and Physics of Lipids* **122**, 97-105.

Anderluh, G., Barlic, A., Potrich, C., Macek, P. & Menestrina, G. 2000a Lysine 77 is a key residue in aggregation of equinatoxin II, a pore-forming toxin from sea anemone *Actinia eguina*. *Journal of Membrane Biology* **173**, 47-55.

Anderluh, G. & Lakey, J. H. 2008 Disparate proteins use similar architectures to damage membranes. *Trends in Biochemical Sciences* **33**, 482-490.

Anderluh, G., Podlesek, Z. & Macek, P. 2000b A common motif in proparts of Cnidarian toxins and nematocyst collagens and its putative role. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1476**, 372-376.

Anderluh, G., Pungercar, J., Strukelj, B., Macek, P. & Gubensek, F. 1995 The Coding Region of the Equinatoxin-Ii Gene Lacks Introns. *Croatica Chemica Acta* **68**, 533-542.

Anderson, P. A. V., Holman, M. A. & Greenberg, R. M. 1993 Deduced Amino-Acid-Sequence of a Putative Sodium-Channel from the Scyphozoan Jellyfish

Cyanea capillata. *Proceedings of the National Academy of Sciences USA* **90**, 7419-7423.

Aneiros, A. & Garateix, A. 2004 Bioactive peptides from marine sources: pharmacological properties and isolation procedures *Journal of Chromatography B* **803**, 41-53.

Antuch, W., Berndt, K. D., Chavez, M. A., Delfin, J. & Wuthrich, K. 1993 The NMR solution structure of a Kunitz-type proteinase inhibitor from the sea anemone *Stichodactyla helianthus*. *European Journal of Biochemistry* **212**.

Assakura, M. T., Silva, A. C., Mentele, R., Camargo, C. M. A. & Serrano, M. T. S. 2003 Molecular cloning and expression of structural domains of bothropasin, a P-III metalloproteinase from the venom of *Bothrops jararaca*. *Toxicon* **41**, 217-227.

Ávila-Soria, G. 2001 Purification parcial y caracterizacion de la actividad biologica de la escifomedusa *Aurelia aurita* del Caribe Mexicano. In *Facultad de Ciencias, Biologia*, vol. Licenciatura, pp. 30. Mexico, D.F: Universidad Nacional Autonoma de México.

Azila, N., Siao, F. K. & Othman, I. 1991 Hemolytic, Edema and Hemorrhage Inducing Activities of Tentacular Extract of the Blubber Jellyfish (Catostylus-Mosaicus). *Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology* **99**, 153-156.

Bachrach, G., Banai, M., Fishman, Y. & Bercovier, H. 1997 Delayed-type hypersensitivity activity of the Brucella L7/L12 ribosomal protein depends on posttranslational modification. *Infection and Immunity* **65**, 267-271.

Balla, G., Jacob, H. S., Balla, J., Rosenberg, m., Nath, k., Apple, F., Eaton, W. J. & Vercellotti, G. M. 1992 Ferritin: a cytoprotective antioxidant strategem of endothelium. *Journal of Biolological Chem*istry **267**, 18148-18153.

Barnes, J. M. 1964 Cause and effect in Irukandji stingings. *Medical Journal of Australia*, 897-904.

Barrett, A. J. 2004 Bioinformatics of proteases in the MEROPS database. *Current Opinion in Drug Discovery & Development* **7**, 334-341.

Bartoli, L., Calabrese, R., Fariselli, F., Mita, D. G. & Casadio, R. 2006 A computational approach for detecting peptidases and their specific inhibitors at the genome level. *MBMC Bioinformatics* **8**, 8.

Batista, I.F., Chudzinski-Tavassi, A.M., Faria, F., Simons, S.M., Barros-Batestti, D. M., Labruna, M.B., Leão, L., H, P & Junqueira-de-Azevedo, I.L. 2008 Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari Ixodidae). *Toxicon* **15**, 823-834.

Bendtsen, J. D., Jensen, l. J., Blom, N., Von Heijne, G. & Brunak, S. 2004b Feature based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection* **17**, 349-356.

Blanchart, A., Fouville, D., Esquibet, M., Mugniery, D. & Grenier, E. 2007 Sequence polymorphism of 2 pioner genes expressed in phytoparasitic nematodes showing different host ranges. *Journal of Heredity*. Advance Access published. DOI:10.1093/jhered/esm050.

Blom, N., Gammeltoft, S. & Brunak, S. 1999 Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* **294**, 1351-1362.

Bloom, D. A., Burnett, J. W. & Alderslade, P. 1998 Partial purification of box jellyfish *Chironex fleckeri* nematocyst venom isolated at the beachside. *Toxicon* **36**, 1075-1085.

Bloom, D. A., Radwan, F. F. Y. & Burnett, J. W. 2001 Toxinological and immunological studies of capillary electrophoresis fractionated *Chrysaora quinquecirrha* fishing tentacle and *Chironex fleckeri* Southcott nematocyst venoms. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* **128**, 75-90.

Bonaldo, M. F., Lennon, G. & Soares, M. B. 1996 Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research* **6**, 791-806.

Bond, J. S. & Beynon, R. J. 1995 The astacin family of metalloendopeptidases. *Protein Science* **4**, 1247-1261.

Bottger, A. & Alexandrova, O. 2007 Programmed cell death in Hydra. *Seminars in Cancer Biology* **17**, 134-146.

Bouchard, C., Ribeiro, P., Dube, F., Demers, C. & Anctil, M. 2004 Identification of a novel aminergic-like G protein-coupled receptor in the cnidarian *Renilla koellikeri*. *Gene* **341**, 67-75.

Bridge, D., Cunningham, C. W., Schierwater, B., Desalle, R. & Buss, L. W. 1992 Class-Level Relationships in the Phylum Cnidaria - Evidence from Mitochondrial Genome Structure. *Proceedings of the National Academy of Sciences USA* **89**, 8750-8753.

Brinkman, D. & Burnell, J. 2007 Identification, cloning and sequencing of two major venom proteins from the box jellyfish, *Chironex fleckeri*. *Toxicon* **50**, 850-860.

Brinkman, D. & Burnell, J. 2008 Partial purification of cytolytic venom proteins from the box jellyfish, *Chironex fleckeri. Toxicon* **51**, 853-863.

Brinkman, L. D. & Burnell, N. J. 2009 Biochemical and Molecular characterisation of cubozoan protein toxins. *Toxicon* **Article in Press, Corrected Proof . doi:10.1016/j.toxicon.2009.02.006**

Burnett, J. W., Calton, G. J. & Burnett, H. W. 1986 Jellyfish Envenomation Syndromes. *Journal of the American Academy of Dermatology* **14**, 100-106.

Burnett, J. W., Long, K. O. & Rubinstein, H. M. 1992 Beachside Preparation of Jellyfish Nematocyst Tentacles. *Toxicon* **30**, 794-796.

Burnett, J. W., Weinrich, D., Williamson, J. A., Fenner, P. J., Lutz, L. L. & Bloom, D. A. 1998 Autonomic neurotoxicity of jellyfish and marine animal venoms. *Clinical Autonomic Research* **8**, 125-130.

Cadot, P., Diaz, J. F., Proost, P., Van Damme, J., Engelborghs, Y., Stevens, E. A. M. & Ceuppens, J. L. 2000 Purification and characterization of an 18-kd allergen of birch (*Betula verrucosa*) pollen: identification as a cyclophilin. *Journal of Allergy and Clinical Immunology* **105**, 286–291. .

Calton, G. J. & Burnett, J. W. 1982 Partial purification ond characterization of the alkaline protease of the sea nettle (*Chrysaora quinquecirrha*) nematocyst venom. *Comparative Biochemistry and Physiology-Physiology* **74**, 361-364.

Carrette, T. & Seymour, J. 2004 A rapid and repeatable method for venom extraction from Cubozoan nematocysts. *Toxicon* **44**, 135-139.

Carrette, T. & Seymour, J. 2006 Cardiotoxic effects of venoms from *Chironex fleckeri* and *Chiropsalmus sp*. on an invertebrate model. *Journal of Venomous Animals and Toxins Including Tropical Diseases* **12**, 245-254.

Cartwright, P., Schierwater, B. & Buss, L. W. 2006 Expression of a Gsx parahox gene, Cnox-2, in colony ontogeny in Hydractinia (Cnidaria : Hydrozoa). *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* **306B**, 460-469.

Castaneda, O., Sotolongo, V., Amor, A. M., Stocklin, R., Anderson, A. J., Harvey, A. L., Engstrom, A., Wernstedt, C. & Karlsson, E. 1995 Characterization of a Potassium Channel Toxin from the Caribbean Sea-Anemone *Stichodactyla-Helianthus*. *Toxicon* **33**, 603-613.

Ceroni, A., Passerini, A., Vullo, A. & Frasconi, P. 2006 DISULFIND: a Disulfide Bonding State and Cysteine Connectivity Prediction Server. *Nucleic Acids Research* **34**.

Cheng, A. C., Winkel, K. D., Hawdon, G. M. & McDonald, M. 1999 Irukandji Like syndrome in Victoria. *Australian and New Zeland Journal of Medicine* **29**, 835.

Chera, S., de Rosa, R., Miljkovic-Licina, M., Dobretz, K., Ghila, L., Kaloulis, K. & Galliot, B. 2006 Silencing of the hydra serine protease inhibitor Kazal1 gene mimics the human SPINK1 pancreatic phenotype. *Journal of Cell Science* **119**, 846-857.

Chung, J. J., Ratnapala, L. A., Cooke, I. M. & Yanagihara, A. A. 2001 Partial purification and characterization of a hemolysin (CAH1) from Hawaiian box jellyfish (*Carybdea alata*) venom. *Toxicon* **39**, 981-990.

Cole, C. N., Barber, J. D. & Barton, G. J. 2008 The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **36**, W197-W201.

Collins, A. G., Schuchert, P., Marques, A. C., Jankowski, T., Medina, M. & Schierwater, B. 2006 Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology* **55**, 97-115.

Cotton, J., Crest, M., Bouet, F., Alessandri, N., Gola, M., Forest, E., Karlsson, E., Castaneda, O., Harvey, A. L., Vita, C. & Menez, A. 1997 A potassium-channel toxin from the sea anemone *Bunodosoma granulifera*, an inhibitor for Kv1 channels - Revision of the amino acid sequence, disulfide-bridge assignment, chemical synthesis, and biological activity. *European Journal of Biochemistry* **244**, 192-202.

Currie, B. J. 2000 Clinical toxicology: A tropical Australian perspective. *Therapeutic Drug Monitoring* **22**, 73-78.

Currie, B. J. 2003 Marine antivenoms. *Journal of Toxicology-Clinical Toxicology* **41**, 301-308.

Daly, M., Brugler, M. R., Cartwright, P., Collins, A. G., Dawson, A. H., Fautin, D. G., France, S. C., McFadden, C. S., Dennis, M., Rodriguez, E., Romano, S. L. & Stake, J. L. 2007 The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years. *Zootaxa* 1668, 127-182.

Darling, J. A., Reitzel, A. R., Burton, P. M., Mazza, M. E., Ryan, J. F., Sullivan, J. C. & Finnerty, J. R. 2005 Rising starlet: the starlet sea anemone, *Nematostella vectensis*. *Bioessays* **27**, 211-221.

Dash, C., Kulkarni, A., Dunn, B. & Rao, M. 2003 Aspartic peptidase inhibitors: implications in drug development. *Biochemistry and Molecular Biology.* **38**, 89-119.

Dauplais, M., Lecoq, A., Song, J. X., Cotton, J., Jamin, N., Gilquin, B., Roumestand, C., Vita, C., deMedeiros, C. L. C., Rowan, E. G., Harvey, A. L. & Menez, A.

1997 On the convergent evolution of animal toxins - Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *Journal of Biological Chemistry* **272**, 4302-4309.

David, C. N., Ozbek, S., Adamczyk, P., Meier, S., Pauly, B., Chapman, J., Hwang, J. S., Gojobori, T. & Holstein, T. W. 2008 Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trend in Genetics* **24**, 431-438.

David, C. N., Schmidt, N., Schade, M., Pauly, B., Alexandrova, O. & Bottger, A. 2005 Hydra and the evolution of apoptosis. *Integrative and Comparative Biology* **45**, 631-638.

De Jong, D. M., Hislop, N. R., Hayward, D. C., Reece-Hoyes, J. S., Pontynen, P. C., Ball, E. E. & Miller, D. J. 2006 Components of both major axial patterning systems of the Bilateria are differentially expressed along the primary axis of a 'radiate' animal, the anthozoan cnidarian *Acropora millepora*. *Developmental Biology* **298**, 632-643.

De Pender, A. M. G., Winkel, K. D. & Ligthelm, R. J. 2006 A probable case of Irukandji syndrome in Thailand. *Journal of Travel Medicine* **13**, 240-243.

Delfin, J., Gonzalez, Y., Diaz, J. & Chavez, M. 1994 Proteinase inhibitor from *Stichodactyla helianthus*: purification, characterization and immobilization, . *Archives of Medical Research*, 199–20. .

Denker, E., Manuel, M., Leclere, L., Le Guyader, H. & Rabet, N. 2008 Ordered progression of nematogenesis from stem cells through differentiation stages in the tentacle bulb of Clytia hemisphaerica (Hydrozoa, Cnidaria). *Developmental Biology* **315**, 99-113.

Dishaw, L., Smith, S. & Bigger, C. 2005 Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics* **57**, 535-548.

Docherty, K. & Steiner, D. F. 1982 Post-translational proteolysis in polypeptide hormone biosyntheis. *Annual Reviews Physiology* **44**, 625-638.

Duckert, P., Brunak, S. & Blom, N. 2004 Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection* **17**, 107-112.

Endean, R., Monks, S. A. & Cameron, A. M. 1993 Toxins from the Box-Jellyfish *Chironex fleckeri*. *Toxicon* **31**, 397-410.

Endean, R. & Sizemore, D. J. 1988 The effectiveness of antivenom in countering the actions of box-jellyfish (*Chironex fleckeri*) nematocyst toxins in mice. *Toxicon* **26**, 425.

Engel, U., Oezbek, S., Engel, R., Petri, B., Lottspeich, F. & Holstein, T. W. 2002 Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in Hydra. *Journal of Cell Science* **115**, 3923-3934.

Engel, U., Pertz, O., Fauser, C., Engel, J., David, C. N. & Holstein, T. W. 2001 A switch in disulfide linkage during minicollagen assembly in Hydra nematocysts. The *EMBO Journal* **20**, 3063-3073.

Enghild, J. J., Salvesen, G., Thøgersen, I. B. & Pizzo, S. V. 1989 Proteinase binding and inhibition by the monomeric alpha-macroglobulin rat alpha 1-inhibitor-3. *Journal of Biological Chemistry* **264**, 11428-35.

Epsztejn, S., Glickstein, H., Picard, V., Slotki, I. N., Breuer, W., Beaumont , C. & Cabantchik, Z. I. 1999 H-ferritin subunit overexpression in erythroid cells reduces the oxidative stress response and induces multidrug resistance properties. *Blood* **94**, 3593-3603.

Estrada-Muñoz, N. A. 2001 Caracterización biológica de las toxinas de *Carybdea marsupialis*. Lineé, 1758 (Cnidaria: Cubozoa). In *Biologia Marina*, vol.

Licenciatura, pp. 153. La Paz, Baja California Sur: Universidad Autonoma de Baja California Sur. México.

Fang, K., Vitale M & Fehlner P, a. T. P. K. 1988 cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5. *Proceeding of the National Academy of Sciences USA.* **85**, 895-899.

Fenner, J. 2006 Jellyfish responsible for Irukandji syndrome. *Qjm-an International Journal of Medicine* **99**, 802-803.

Fenner, P. J. & Hadok, J. C. 2002 Fatal envenomation by jellyfish causing Irukandji syndrome. *Medical Journal of Australia* **177**, 362-363.

Fenner, P. J. & Hadok, J. C. 2003 Fatal envenomation by jellyfish causing Irukandji syndrome. *Medical Journal of Australia* **178**, 139-140.

Fenner, P., Rodgers, D. & Williamson, J. 1986a Box Jellyfish Antivenom and Irukandji Stings. *Medical Journal of Australia* **144**, 665-666.

Fenner, P. J., Williamson, J., Callanan, V. I. & Audley, I. 1986b Further understanding of, and a new treatment for Irukandji (*Carukia barnesi*) stings *Medical Journal of Australia* **145**.

Finnerty, J. R. & Martindale, M. Q. 1999 Early evolution of Hox and ParaHox genes: Evidence from the Cnidaria. *Developmental Biology* **210**, 187-187.

Flecker, H. 1952 Irukandji sting to north Queensland bathers without of weals but with severe general symtoms. *Medical Journal of Australia*, 89-91.

Flueckiger, S., Fijten, H., Whitley, P., Blaser, K. & Crameri, R. 2002 Cyclophilins, a new family of cross-reactive allergens. *European Journal of Immunology* **32**, 10-17.

Galliot, B. & Miller, D. 2000 Origin of anterior patterning - how old is our head? *Trends in Genetics* **16**, 1-5.

Garm, A., Coates, M. M., Gad, R., Seymour, J. & Nilsson, D. E. 2007a The lens eyes of the box jellyfish *Tripedalia cystophora* and *Chiropsalmus sp.* are slow and color-blind. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology* **193**, 547-557.

Garm, A., Poussart, Y., Parkefelt, L., Ekstrom, P. & Nilsson, D. E. 2007b The ring nerve of the box jellyfish *Tripedalia cystophora*. *Cell and Tissue Research* **329**, 147-157.

Gasparini, S., Gilquin, B. & Menez, A. 2004 Comparison of sea anemone and scorpion toxins binding to Kv1 channels: an example of convergent evolution. *Toxicon* **43**, 901-908.

Gawlik, K. & Poreba, W. G. 2005 Cystatins, thyropins and inhibitors homologous to propeptides of cysteine proteases. *Postepy Biochemii.* **51**, 318-327.

Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. 2002 protein homology by domain architecture. *Genome Research* **12**, 1619–1623.

Gendeh, G. S., Chung, M. C. M. & Jeyaseelan, K. 1997a Genomic structure of a potassium channel toxin from *Heteractis magnifica*. *FEBS Letters* **418**, 183-188.

Gendeh, G. S., Young, L. C., De Medeiros, C. L. C., Jeyaseelan, K., Harvey, A. L. & Chung, M. C. M. 1997b A New Potassium Channel Toxin from the Sea Anemone *Heteractis magnifica*: Isolation, cDNA Cloning, and Functional Expression. *Biochemistry* **36**, 11461-11471.

George Shackelford, K. K. 2007 Contact prediction using mutual information and neural nets. *Proteins: Structure, Function, and Bioinformatics* **69**, 159-164.

Gerhart, D. J. 1991 Emesis, Learned Aversion, and Chemical Defense in Octocorals - a Central Role for Prostaglandins. *American Journal of Physiology* **260**, R839-R843.

Gershwin, L. 2005a Taxonomy and phylogeny of Australia Cubozoa. In *School of Marine Biology and Aquaculture.*, vol. PhD, pp. 221., 49 plates. Townsville, Australia: James Cook University.

Gershwin, L. 2006a Jellyfish responsible for Irukandji syndrome. *Qjm-an International Journal of Medicine* **99**, 801-802.

Gershwin, L. A. 2005b Two new species of jellyfishes (Cnidaria : Cubozoa : Carybdeida) from tropical western Australia, presumed to cause Irukandji syndrome. *Zootaxa*, 1-30.

Gershwin, L. A. 2006b Nematocysts of the Cubozoa. *Zootaxa*, 1-57.

Gershwin, L. A. 2007a *Malo kingi*: A new species of Irukandji jellyfish (Cnidaria : Cubozoa : Carybdeida), possibly lethal to humans, from Queensland, Australia. *Zootaxa*, 55-68.

Gershwin, L. A. 2007b *Malo kingi*: A new species of Irukandji jellyfish (Cnidaria: Cubozoa: Carydeida), possibly lethal to humans, from Queensland, Australia. *Zootaxa* **1659**.

Gilquin, B., Braud, S., Eriksson, M. A. L., Roux, B., Bailey, T. D., Priest, B. T., Garcia, M. L., Menez, A. & Gasparini, S. 2005 A variable residue in the pore of Kv1 channels is critical for the high affinity of Blockers from sea anemones and scorpions. *Journal of Biological Chemistry* **280**, 27093-27102.

Gish, W. & States, D. J. 1993 Identification of Protein Coding Regions by Database Similarity Search. *Nature Genetics* **3**, 266-272.

Gough, J. a. K., K and Hughey, R and Chothia, C. 2001 Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology* **313**, 903-919.

Grady, J. D. & Burnett, J. W. 2003 Irukandji-like syndrome in South Florida divers. *Annals of Emergency Medicine* **42**, 763-766.

Guder, C., Pinho, S., Nacak, T. G., Schmidt, H. A., Hobmayer, B., Niehrs, C. & Holstein, T. W. 2006 An ancient Wnt-dickkopf antagonism in Hydra. *Development* **133**, 901-911.

Guex, N. & Peitsch, M. C. 1997 SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

Gusmani, L., Avian, M., Galil, B., Patriarca, P. & Rottini, G. 1997 Biologically active polypeptides in the venom of the jellyfish Rhopilema nomadica. *Toxicon* **35**, 637-648.

Habetha, M. & Bosch, T. 2005 Symbiotic Hydra express a plant-like peroxidase gene during oogenesis. *Journal of Experimental Biology* **208**, 2157-2164.

Hana, J., Youa, D., Xua, X., Hanc, W., Lub, Y., Laia, R. & Menga, Q. 2008 An anticoagulant serine protease from the wasp venom of *Vespa magnifica*. *Toxicon* **51**, 914-922.

Hargreaves, D. C. & Medzhitov, R. 2005 Innate sensors of microbial insects. *Journal of Clinical Immunology* **25**.

Hasegawa, Y., Honma, T., Nagai, H., Ishida, M., Nagashima, Y. & Shiomi, K. 2006 Isolation and cDNA cloning of a potassium channel peptide toxin from the sea anemone *Anemonia erythraea*. *Toxicon* **48**, 536-542.

Hayakawa H, A. T., Watanabe T. 2002 recursor structure of egg proteins in the coral *Galaxea fascicularis. Biochemical and Biophysical Research Communications* **344**.

Hayward, D. C., Catmull, J., Reece-Hoyes, J. S., Berghammer, H., Dodd, H., Hann, S. J., Miller, D. J. & Ball, E. E. 2001 Gene structure and larval expression of cnox-2Am from the coral *Acropora millepora*. *Development Genes and Evolution* **211**, 10-19.

Hellstern, S., Stetefeld, J., Fauser, C., Lustig, A., Engel, J., Holstein, T. W. & Ozbek, S. 2006 Structure/function analysis of spinalin, a spine protein of Hydra nematocysts. *FEBS Journal* **273**, 3230-3237.

Hemmrich, G., Miller, D. J. & Bosch, T. C. G. 2007 The evolution of immunity: a low-life perspective. *Trends in Immunology* **28**, 449-454.

Heralde, F. M. H., Imperial, j., Bandyopadhyayc, P. K., Olivera, B. M., Concepcion, J. P. & Santosa, A. D. 2008 A rapidly diverging superfamily of peptide toxins in venomous Gemmula species. *Toxicon* **51**, 890-897.

Herzig, V., Ward, R. J. & Ferreira dos Santos, W. 2004 Ontogenetic changes in Phoneutria nigriventer (Araneae, Ctenidae) spider venom *Toxicon* **44**, 635-640.

Hofmann, K. & Stoffel, W. 1993 TMbase - A database of membrane spanning proteins segments. *Biological Chemistry Hoppe-Seyler* **374**, 166.

Holstein, T. W., Benoit, M., Vonherder, G., Wanner, G., David, C. N. & Gaub, H. E. 1994 Fibrous Mini-Collagens in Hydra Nematocysts. *Science* **265**, 402-404.

Holstein, T. W., Hobmayer, E. & Technau, U. 2003 Cnidarians: An evolutionarily conserved model system for regeneration? *Developmental Dynamics* **226**, 257-267.

Honma, T., Hasegawa, Y., Ishida, M., Nagai, H., Nagashima, Y. & Shiomi, K. 2005a Isolation and molecular cloning of novel peptide toxins from the sea anemone *Antheopsis maculata*. *Toxicon* **45**, 33-41.

Honma, T., Minagawa, S., Nagai, H., Ishida, M., Nagashima, Y. & Shiomi, K. 2005b Novel peptide toxins from acrorhagi, aggressive organs of the sea anemone *Actinia equina*. *Toxicon* **46**, 768-774.

Hudson, J. R., Bridge, N. F., Walker, T. I. & Victoria. Marine and Freshwater Resources Institute. 1997 Feasibility study for establishment of a Victorian commercial jellyfish fishery. Queenscliff, Vic. *Marine and Freshwater Resources Institute.*

Huynh, T. T., Seymour, J., Pereira, P., Mulcahy, R., Cullen, P., Carrette, T. & Little, M. 2003 Severity of Irukandji syndrome and nematocyst identification from skin scrapings. *Medical Journal of Australia* **178**, 38-41.

Iguchi, A., Iwanaga, S. & Nagai, H. 2008 Isolation and characterization of a novel protein toxin from fire coral. *Biochemical and Biophysical Research Communications* **365**, 107-112.

Imler, J. L. & Hoffman, J. A. 2000 Signaling mechansims in the antimicrobial host defense of Drosophila. *Current Opinion in Microbiology* **3**, 16-22.

Jensen, L. J., Gupta, R., Blom, N., Devos, D., Tamames, J., Kesmir, C., Nielsen, H., Stærfeldt, H., Workman, C., Andersen, C. A. F., Knudsen, S., Krogh, A., Valencia, A. & Brunak, S. 2002 *Ab initio* prediction of human orphan protein function from post-translational modifications and localization features. *Journal of Molecular Biology* **319**, 1257-1265.

Jensen, L. J., Stærfeldt, H.-H. & Søren, B. 2003 Prediction of human protein function according to Gene Ontology categories. *Bioinformatics*, **19**, 635-642.

Jeziorski, M. C., Greenberg, R. M., Clark, K. S. & Anderson, P. A. V. 1998 Cloning and functional expression of a voltage-gated calcium channel alpha(1) subunit from jellyfish. *Journal of Biological Chemistry* **273**, 22792-22799.

Jones, D., Sawicki, G. & Wozniak, M. 1992 Sequence Structure, and Expression of a Wasp Venom Protein with a Negatively Charged Signal Peptide and Novel Repeat Internal Structure. *The Journal of Biological Chemistry* **267**, 14871-14878.

Jones, D. T. 1999 Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* **292**, 195–202.

Julenius, K., Mølgaard, A., Gupta, R. & Brunak, S. 2005 Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* **15**, 153-164.

Kath, J. C., Hanson, D. C. & Chandy, K. G. 1997 T lymphocyte potassium channel blockers. *Annual Reports in Medicinal Chemistry, Vol 32* **32**, 181-190.

Katzman, S., Barrett, C., Thiltgen, G., Karchin, R. & Karplus, K. 2008 PREDICT-2ND: a tool for generalized protein local structure prediction. *Bioinformatics* **24**, 2453-2459.

Kem, W. R., Sanyal, G., Williams, R. W. & Pennington, M. W. 1996 Secondary structure of ShK toxin, a potassium-channel-blocking peptide. *Letters in Peptide Science* **3**, 69-72.

Khoo, K. S., Kam, W. K., Khoo, H. E., Gopalakrishnakone, P. & Chung, M. C. M. 1993 Purification and Partial Characterization of 2 Cytolysins from a Tropical Sea Anemone, *Heteractis-Magnifica*. *Toxicon* **31**, 1567-1579.

Kinsey, B., Barnes, J. T. & Sir George Fisher Centre for Tropical Marine Studies. 1988 More Barnes on box jellyfish. Townsville, Qld.: Sir George Fisher Centre for Tropical Marine Studies, James Cook University of North Queensland.

Kitajima, S. & Sato, F. 1999 Plant Pathogenesis-Related Proteins: Molecular Mechanisms of Gene Expression and Protein Function. *Journal of Biochemistry* **125**, 1-8.

Knudsen, A. & Vedeler, C. A. 2006 cDNA expression library screening for identoficastion of novel onconeural antigens. *Acta Neurologica Scandinovica* **113**, 73-74.

Kortschak, R. D., Samuel, G., Saint, R. & Miller, D. J. 2003 EST analysis of the Cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Current Biology* **13**, 2190-2195.

Kozak, M. 1989 The scanning model for translation: an update. *Journal of Cell Biology* **108**, 229–241.

Kozlov, S. A. & Grishin, E. V. 2007 The universal algorithm of maturation for secretory and excretory protein precursors. *Toxicon* **49**, 721-726.

Kozmik, Z., Swamynathan, S. K., Ruzickova, J., Jonasova, K., Paces, V., Vlcek, C. & Piatigorsky, J. 2008 Cubozoan crystallins: evidence for convergent evolution of pax regulatory sequences. *Evolution & Development* **10**, 52-61.

Kristan, K., Viero, G., Macek, P., Dalla Serra, M. & Anderluh, G. 2007 The equinatoxin N-terminus is transferred across planar lipid membranes and helps to stabilize the transmembrane pore. *FEBS Journal* **274**, 539-550.

Lanio, M. E., Morera, V., Alvarez, C., Tejuca, M., Gomez, T., Pazos, F., Besada, V., Martinez, D., Huerta, V., Padron, G. & Chavez, M. D. 2001 Purification and characterization of two hemolysins from Stichodactyla helianthus. *Toxicon* **39**, 187-194.

Leggat, W., Hoegh-Guldberg, O., Dove, S. & Yellowlees, D. 2007 Analysis of an EST library from the dinoflagellate (*Symbiodinium sp.*) symbiont of reef-building corals. *Journal of Phycology* **43**, 1010-1021.

Lindgens, D., Holstein, T. W. & Technau, U. 2004 Hyzic, the Hydra homolog of the zic/odd-paired gene, is involved in the early specification of the sensory nematocytes. *Development* **131**, 191-201.

Liping J, Li P, Jinjun C, Yongqun Z, Xia X & L, S. 2008 Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*. *Toxicon* 51(8):1479-89

Little, M. & Mulcahy, R. F. 1998 A year's experience of Irukandji envenomation in far north Queensland. *Medical Journal of Australia* **169**, 638-641.

Little, M., Pereira, P., Carrette, T. & Seymour, J. 2006 Jellyfish responsible for Irukandji syndrome. *Qjm-an International Journal of Medicine* **99**, 425-427.

Lotan, A., Fishman, L. & Zlotkin, E. 1996 Toxin compartmentation and delivery in the cnidaria: The nematocyst's tubule as a multiheaded poisonous arrow. *Journal of Experimental Zoology* **275**, 444-451.

Macek, P., Belmonte, G., Pederzolli, C. & Menestrina, G. 1994 Mechanism of Action of Equinatoxin-Ii, a Cytolysin from the Sea-Anemone *Actinia-Equina* L Belonging to the Family of Actinoporins. *Toxicology* **87**, 205-227.

Mackessy, S. P., Sixberry, N. M., Heyborne, W. H. & Fritts, T. 2006 Venom of the Brown Treesnake, Boiga irregularis: Ontogenetic shifts and taxa-specific toxicity *Toxicon* **47**, 537-548.

Malej, A. & Malej, A. 1987 Jellyfish *Pelagia Noctiluca* and Medical Significance of Its Sting. *Zdravstveni Vestnik* **56**, 161-163.

Malovrh, P., Viero, G., Dalla Serra, M., Podlesek, Z., Lakey, J. H., Macek, P., Menestrina, G. & Anderluh, G. 2003 A novel mechanism of pore formation - Membrane penetration by the N-terminal amphipathic region of equinatoxin. *Journal of Biological Chemistry* **278**, 22678-22685.

Mans, B. J., Andersen, J. F., Francischetti, I. M., Valenzuela, J. G., Schwan, T. G., Pham, V. M., Garfield, M. K., Hammer, C. H. & Ribeiro, J. M. 2008 Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behaviour Insect. *Biochemistry and Molecular. Biology*. **38**, 42-58.

Marchini, B., De Nuccio, L., Mazzei, M. & Mariottini, G. L. 2004 A fast centrifuge method for nematocyst isolation from *Pelagia noctiluca Forskal* (Cnidaria : Scyphozoa). *Rivista Di Biologia Biology Forum* **97**, 505-515.

Mariottini, G. L., Sottofattori, E., Mazzei, M., Robbiano, L. & Carli, A. 2002 Cytotoxicity of the venom of *Pelagia noctiluca forskal* (Cnidaria : Scyphozoa). *Toxicon* **40**, 695-698.

Marsh, L. M., Slack-Smith, S. M. & Gurry, D. L. 1986a *Sea stingers : and other venomous and poisonous marine invertebrates of Western Australia*. Perth, W.A.: Western Australian Museum.

Marsh, L. M., Slack-Smith, S. M., Gurry, D. L. & Western Australian Museum. 1986b *Sea stingers: and other venomous and poisonous marine invertebrates of Western Australia*. Perth, W.A.: Western Australian Museum.

Martin-Eauclaire, M.-F., SoGaard, M., Ramos, C., Cestele, S., Bougis, P. E. & Svensson, B. 1994 Production of active, insect-specific scorpion neurotoxin in yeast. *European Journal of Biochemistry* **223**, 637-645.

Matveev, I. V., Shaposhnikova, T. G. & Podgornaya, O. I. 2007 A novel Aurelia aurita protein mesoglein contains DSL and ZP domains. *Gene* **399**, 20-25.

Mayer, C., Appenzeller, U., Seelbach, H., Achatz, G., Oberkofler, H., Breitenbach, M., Blaser, K. & Crameri, R. 1999 Humoral and Cell-mediated Autoimmune Reactions to Human Acidic Ribosomal P2 Protein in Individuals Sensitized to *Aspergillus fumigatus* P2 Protein. *Journal of Experimental Medicine* **189**, 1507-1512.

McInerney, J. O. 1998 GCUA (General Codon Usage Analysis). *Bioinformatics* **14**, 372-373.

Messerli, S. M. & Greenberg, R. M. 2006 Cnidarian toxins acting on voltage-gated ion channels. *Marine Drugs* **4**, 70-81.

Miljkovic-Licina, M., Gauchat, D. & Galliot, B. 2004 Neuronal evolution: analysis of regulatory genes in a first-evolved nervous system, the hydra nervous system. *Biosystems* **76**, 75-87.

Miller, D. J., Ball, E. E. & Technau, U. 2005 Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends in Genetics* **21**, 536.

Miller, D. J., Hemmrich, G., Ball, E. E., Hayward, D. C., Khalturin, K., Funayama, N., Agata, K. & Bosch, T. C. G. 2007 The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. *Genome Biology* **8**, -.

Minagawa, S., Ishida, M., Nagashima, Y. & Shiomi, K. 1998 Primary structure of a potassium channel toxin from the sea anemone *Actinia equina*. *FEBS Letters* **427**, 149-151.

Mittermann, I., Voronin, V., Heberle-Bors, E. & Valenta, R. 2005 Identification of a villin-related tobacco protein ad a novel cross-reactive plant allergens. *FEBS Letters.*

Mohrlen, F., Hutter, H. & Zwilling, R. 2003 The astacin protein family in *Caenorhabditis elegans. European Journal of Biochemistry* **270**, 4909-4920.

Mohrlen, F., Maniura, M., Plickert, G., Frohme, M. & Frank, U. 2006 Evolution of astacin-like metalloproteases in animals and their function in development. *Evolution & Development* **8**, 223-231.

Morales-Landa, J. L., Zapata-Perez, O., Cedillo-Rivera, R., Segura-Puertas, L., Sima-Alvarez, R. & Sanchez-Rodriguez, J. 2007 Antimicrobial, antiprotozoal, and toxic activities of cnidarian extracts from the Mexican Caribbean sea. *Pharmaceutical Biology* **45**, 37-43.

Moran, Y. & Gurevitz, M. 2006 When positive selection of neurotoxin genes is missing. The riddle of the sea anemone *Nematostella vectensis. FEBS Journal* **273**, 3886-3892.

Moran, Y., Weinberger, H., Reitzel, A. M., Sullivan, J. C., Kahn, R., Gordon, D., R., F. J. & Gurevitz, M. 2008a Intron retention as a posttranscriptional regulatory mechanism of neurotoxin expression at early life stages of the starlet anemone *Nematostella vectensis. Journal of Molecular BIology* **380**, 437-443.

Moran, Y., Weinberger, H., Sullivan, J. C., Reitzel, A. M., Finnerty, J. R. & Gurevitz, M. 2008b Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the *Nematostella vectensis* genome. *Molecular Biology and Evolution* **25**, 737-747.

Morris, S. C. 1993 The fossil record and the early evolution of the Metazoa. *Nature* **136**, 219-225.

Moussoukhoye, S. D. & Abdoulaye, S. 2004 Identification de glycolipides isolés d'algues et de cnidaire de la côte sénégalaise. *C. R. Chimie* **7**, 956-971.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. 1995 SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* **247**, 536-540.

Muta, T. & Iwanaga, S. 1996 The role of hemolymph coagulation in innate immunity. *Current Opinion in Immunology* **8,** 41-47.

Nagai, H. 2003 Recent progress in jellyfish toxin study. *Journal of Health Science* **49**, 337-340.

Nagai, H., Takuwa, K., Nakao, M., Ito, E., Miyake, M., Noda, M. & Nakajima, T. 2000a Novel proteinaceous toxins from the box jellyfish (sea wasp) *Carybdea rastoni*. *Biochemical and Biophysical Research Communications* **275**, 582-588.

Nagai, H., Takuwa, K., Nakao, M., Sakamoto, B., Crow, G. L. & Nakajima, T. 2000b Isolation and characterization of a novel protein toxin from the Hawaiian box jellyfish (sea wasp) *Carybdea alata*. *Biochemical and Biophysical Research Communications* **275**, 589-594.

Nagai, H., Takuwa-Kuroda, K., Nakao, M., Oshiro, N., Iwanaga, S. & Nakajima, T. 2002 A novel protein toxin from the deadly box jellyfish (Sea Wasp, Habu-kurage) *Chiropsalmus quadrigatus*. *Bioscience Biotechnology and Biochemistry* **66**, 97-102.

Nagase, H., Karaki, H., Ozaki, H., Urakawa, N., Azuma, H. & Nakajima, T. 1986 Endothelium-Dependent Relaxation Induced by Jellyfish Toxin (Pcrtx) in Rat Aorta. *Japanese Journal of Pharmacology* **40**, P168-P168.

Namshin, K., Seokmin, S. & Sanghyuk, L. 2005 Genome-based EST clustering and gene modeling for alternative splicing. *Genome Res.* **15**, 566-576.

Neeman, I., Calton, G. J. & Burnett, J. W. 1979 Endonucleases from Jellyfish Venoms. *Clinical Research* **27**, A243-A243.

Nevalainen, T. J., Peuravuori, H. J., Quinn, R. J., Llewellyn, L. E., Benzie, J. A. H., Fenner, P. J. & Winkel, K. D. 2004 Phospholipase A2 in Cnidaria. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **139**, 731-735.

Nielsen, H., Engelbrecht, J., Brunak, S. & Heijne, V. G. 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1-6.

Nielsen, H. & Krogh, A. 1998 Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB 6),AAAI Press, Menlo Park, California.*

Noguchi, K., Sakanashi, M., Matsuzaki, T., Nakasone, J., Sakanashi, M., Koyama, T., Hamadate, N. & Sakanashi, M. 2005 Cardiovascular effects and lethality of venom from nematocysts of the box-jellyfish *Chiropsalmus quadrigatus* (Habu-kurage) in anaesthetized rats. *Toxicon* **45**, 519-526.

Nordström, K. 2003 Evolution of eyes: Pax, gene duplications and morphology. In *Department of Cell and Organism Biology*, vol. Ph.D, pp. 120. Lund: Lund University.

O' Reilly, G. M., Isbister, G. K., Lawrie, P. M., Treston, G. T. & Currie, B. J. 2001 Prospective study of jellyfish stings from tropical Australia, including the major box jellyfish *Chironex fleckeri*. *Medical Journal of Australia* **175**, 652-655.

O'Rourke, D., Baban, D., Demidova, M., Richard., M. & Hodgkin, J. 2006 Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. *Genome Research* **10.1101/gr.50823006 doi:** .

Onizuka, R., Kamiya, H., Muramoto, K., Goto, R., Inoue, K., Kumamoto, K., Nakajima, Y., Iida, S. & Ishigami, F. 2001 Purification of the major allergen of red soft coral (*Dendronephthya nipponica*). *International Archives of Allergy and Immunol* **125**, 135-143.

Ovchinnikova, T. V., Balandin, S. V., Aleshina, G. M., Tagaev, A. A., Leonova, Y. F., Krasnodembsky, E. D., Menshenin, A. V. & Kokryakov, V. N. 2006 Aurelin, a novel antimicrobial peptide from jellyfish *Aurelia aurita* with structural features of defensins and channel-blocking toxins. *Biochemical and Biophysical Research Communications* **348**, 514-523.

Ozaki, H., Karaki, H., Nagase, H., Urakawa, N., Azuma, H. & Nakajima, T. 1986 Contractile Effects of Jellyfish Toxin (Pcrtx) on Isolated Vascular Smooth-Muscle of Rabbit Aorta. *Japanese Journal of Pharmacology* **40**, P169-P169.

Page, M. J. & Di Cera, E. 2008 Evolution of peptidase diversity *Journal of Biological Chemistry*.

Pan, T. L., Groger, H., Schmid, V. & Spring, J. 1998 A toxin homology domain in an astacin-like metalloproteinase of the jellyfish *Podocoryne carnea* with a dual role in digestion and development. *Development Genes and Evolution* **208**, 259-266.

Pearse, B. M. 1976 Clathrin: a unique protein associated with intracellular transfer of membrane by coated vesicles. *Proceeding of the National Academy of Sciences USA* **73**, 1255-1259.

Pi, C., Liu, Y. S., Peng, C., Jiang, X. Y., Liu, J., Xu, B., Yu, X., Yu, Y., Jiang, X., Wang, L., Chen, S. & Xu, A. 2006 Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie* **88**, 131-140.

Pommier, P., Coulange, M. & De Haro, L. 2005 Envenimation systemique par meduse en Guadeloupe: Irukandji - like syndrome ? *Medicine Tropicale* **25**, 367-369.

Puce, S., Basile, G., Bavestrello, G., Bruzzone, S., Cerrano, C., Giovine, M., Arillo, A. & Zocchi, E. 2004 Abscisic acid signaling through cyclic ADP-ribose in hydroid regeneration. *Journal of Biological Chemistry* **279**, 39783-39788.

Puente, X. S., Sanchez, L. M., Overall, C. M. & Lopez-Otin, C. 2003 Human and Man proteases: A comparative genomic approach. *Nature Reviews Genetics* **4**, 544-558.

Pung, Y. F., Kumar, S. V., Rajagopalan, N., Fry , B. G., Kumar, P. P. & Kini, R. M. 2006 Ohanin, a novel protein from king cobra venom: Its cDNA and genomic organization *Gene* **371**, 246-256.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., Jurka, J., Genikhovich, G., Grigoriev, I. V., Lucas, S. M., Steele, R. E., Finnerty, J. R., Technau, U., Martindale, M. Q. & Rokhsar, D. S. 2007 Sea anemone genome reveals

ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86-94.

Radwan, F. F. Y., Burnett, J. W., Bloom, D. A., Coliano, T., Eldefrawi, M. E., Erderly, H., Aurelian, L., Torres, M. & Heimer-de la Cotera, E. P. H. D. 2001 A comparison of the toxinological characteristics of two Cassiopea and Aurelia species. *Toxicon* **39**, 245-257.

Radwan, F. F. Y., Gershwin, L. A. & Burnett, J. W. 2000 Toxinological studies on the nematocyst venom of *Chrysaora achlyos*. *Toxicon* **38**, 1581-1591.

Radwan, F. F. Y., Roman, L. G., Baksi, K. & Burnett, J. W. 2005 Toxicity and mAChRs binding activity of *Cassiopea xamachana* venom from Puerto Rican coasts. *Toxicon* **45**, 107-112.

Ramasamy, S., Isbister, G. K., Seymour, J. E. & Hodgson, W. C. 2003 The in vitro effects of two chirodropid (*Chironex fleckeri* and *Chiropsalmus sp*.) venoms: efficacy of box jellyfish antivenom. *Toxicon* **41**, 703-711.

Ramasamy, S., Isbister, G. K., Seymour, J. E. & Hodgson, W. C. 2004 The in vivo cardiovascular effects of box jellyfish *Chironex fleckeri* venom in rats: efficacy of pre-treatment with antivenom, verapamil and magnesium sulphate. *Toxicon* **43**, 685-690.

Ramasamy, S., Isbister, G. K., Seymour, J. E. & Hodgson, W. C. 2005a The in vivo cardiovascular effects of an Australasian box jellyfish (*Chiropsalmus sp*.) venom in rats. *Toxicon* **45**, 321-327.

Ramasamy, S., Isbister, G. K., Seymour, J. E. & Hodgson, W. C. 2005b The in vivo cardiovascular effects of the Irukandji jellyfish (*Carukia barnesi*) nematocyst venom and a tentacle extract in rats. *Toxicology Letters* **155**, 135-141.

Rawlings, N., Tolle, D. P. & Barret, A. J. 2004 Evolutionary families of peptidase inhibitors *Journal of Biochemistry* **378**, 705-716.

Rawlings, N. D. & Barrett, A. J. 1995 Evolutionary families of metallopeptidases. *Methods in Enzymology* **248**, 183-228.

Rawlings, N. D., Morton, F. R. & Barrett, A. J. 2006 MEROPS: the peptidase database *Nucleic Acids Research* **34**, D270-D272.

Rawlings, N. D., O'Brien, E.A. & Barrett, A.J. 2002 MEROPS: the protease database. *Nucleic Acids Research* 343-346

Rawlings, N. R., Morton, F. R., Kok, c. y., Kong, J. & Barrett, A. J. 2008 MEROPS: the peptidase database *Nucleic Acids Research* **36**, D314-D319.

Rojas, A. & Doolittle, R. F. 2002 The occurrence of type S1A serine proteases in sponge and jellyfish. *Journal of Molecular Evolution* **55**, 790-794.

Rost, B. 1996 PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods in Enzymology* **266**.

Rottini, G., Gusmani, L., Parovel, E., Avian, M. & Patriarca, P. 1995 Purification and Properties of a Cytolytic Toxin in Venom of the Jellyfish *Carybdea-Marsupialis*. *Toxicon* **33**, 315-326.

Ruppert, E. E., Fox, S. R. & Barnes, D. R. 2004 *Invertebrate Zoology. A functional evolutionary approach*. United States of America: Thomson, Brooks/Cole.

Saha, S. & Raghava, G. P. S. 2007 Prediction of neurotoxins based on their function and source *In Silico Biology* **7**.

Sambrook, J. a. & Russell, D. W. 2001 *Molecular Cloning: A Laboratory Manual (Third edition)*. Cold Spring Harbor, New York.: Cold Spring Harbor Laboratory Press,.

Sanchez-Rodriguez, J., Torres, E. & Segura-Puertas, L. 2006 Partial purification and characterization of a novel neurotoxin and three cytolysins from box jellyfish (*Carybdea marsupialis*) nematocyst venom. *Archives of Toxicology* **80**, 163-168.

Santos, A. N., Korber, S., Kullertz, G., Fischer, G. & Fischer, B. 2000 Oxygen stress increases prolyl cis/trans isomerase activity and expression of cyclophilin 18 in rabbit blastocysts. *Biology of Reproduction* **62**, 1-7.

Saphire, A. C., Bobardt, M. D. & Gallay, P. A. 1999 Host cyclophilin A mediates HIV-1 attachment to target cells via heparans. *EMBO Journal*, 6771–6785.

Sarras, M. P., Yan, L., Leontovich, A. & Zhang, J. S. 2002 Structure, expression, and developmental function of early divergent forms of metalloproteinases in Hydra. *Cell Research* **12**, 163-176.

Saxena, S., Madan, T., Muralidhar, K. & Sarmp, U. 2003 cDNA cloning, expression and characterization of an allergenic L3 ribosomal protein of *Aspergillus fumigatus. Clinical and Experimental Immunulogy* **134**, 886-91.

Schulte, W., Topfer, R., Stracke, R., Schell, J. & Martini, N. 1997 Multi-functional acetyl-CoA carboxylase from Brassica napus is encoded by a multi-gene family: Indication for plastidic localization of at least one isoform. *Proceedings of the National Academy of Sciences USA* **94**, 3465-3470.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. 1998 SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences USA* **95**, 5857-5864.

Segura-Puertas, L., Ávila-Soria, G., Sanchez-Rodriguez, J & Burnett, J. W. 2002 Some Toxinological Aspects of Aurelia aurita (LINNÉ) from the Mexican Caribbean. *Journal of Venomous Animals and Toxins* **8**, 269-282.

Seipp, S., Wittig, K., Stiening, B., Bottger, A. & Leitz, T. 2006 Metamorphosis of Hydractinia echinata (Cnidaria) is caspase-dependent. *International Journal of Developmental Biology* **50**, 63-70.

Shang, D. & Miller, D. J. 1995 Characterization of a novel non-muscle myosin-related protein from *Onchocerca gibsoni*. *International Journal for Parasitology* **25**, 1385-1391.

Sher, D., Fishman, Y., Zhang, M. L., Lebendiker, M., Gaathon, A., Mancheno, J. M. & Zlotkin, E. 2005a Hydralysins, a new category of beta-pore-forming toxins in Cnidaria. *Journal of Biological Chemistry* **280**, 22847-22855.

Sher, D., Knebel, A., Bsor, T., Nesher, N., Tal, T., Morgenstern, D., Cohen, E., Fishman, Y. & Zlotkin, E. 2005b Toxic polypeptides of the hydra - a bioinformatic approach to cnidarian allomones. *Toxicon* **45**, 865-879.

Shu, Q. & Liang, S. P. 1999 Purification and characterization of huwentoxin-II, a neurotoxic peptide from the venom of the Chinese bird spider *Selenocosmia huwena*. *Journal of Peptide Research* **53**, 486-491.

Shuqing, L., Ming-Zhong, S., Changkai, S., Baochang, Z., Frederick, T. G. & Qingyin, Z. 2008 A novel serine protease from the snake venom of Agkistrodon *blomhoffii ussurensis. Toxicon*
doi: 10.1016/j.toxicon.2008.08.012

Söderhäll, K., Cerenius, l. & Johansson, M. 1996 The prohenoloxidase activating system in invertebrates. In *New Directions in Invertebrate Immunology* (ed. Vasta). New York: SOS

Sottrup-Jensen, L. 1987 *The Plasma Proteins Structure, Funtion and Genetic Control*. New York. Academic Press.

Spagnuolo, A., Zanetti, L., Cariello, L. & Piccoli, R. 1994 Isolation and Characterization of 2 Genes Encoding Calitoxins, Neurotoxic Peptides from *Calliactis-Parasitica* (Cnidaria). *Gene* **138**, 187-191.

Sullivan, J. C. & Finnerty, J. R. 2007 A surprising abundance of human disease genes in a simple "basal" animal, the starlet sea anemone (*Nematostella vectensis*). *Genome* **50**, 689-692.

Sullivan, J. C., Ryan, J. F., Watson, J. A., Webb, J., Mullikin, J. C., Rokhsar, D. & Finnerty, J. R. 2006 StellaBase: The *Nematostella vectensis* Genomics Database. *Nucleic Acids Research* **34**, D495-D499.

Tahirov, T. H., Lu, T. H., Liaw, Y. C., Chen, Y. L. & Lin, J. Y. 1995 Crystal structure of abrin-a at 2.14 A. *Journal of Molecular BIology* **250**, 354-367.

Talvinen, K. A. & Nevalainen, T. J. 2002 Cloning of a novel phospholipase A2 from the cnidarian *Adamsia carciniopados*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **132**, 571-578.

Tamkun, M. N. & Hessinger, D. A. 1981 Isolation and partial characterization of a hemolytic and toxic protein from the nematocyst venom of the Portuguese man-o'-war, Physalia physalis. *Biochemica et Biophysica Acta* 87-98.

Tardent, P. 1995 The Cnidarian Cnidocyte, a High-Tech Cellular Weaponry. *Bioessays* **17**, 351-362.

Tardent, P. & Holstein, T. 1982 Morphology and Morphodynamics of the Stenotele Nematocyst of *Hydra-Attenuata Pall* (Hydrozoa, Cnidaria). *Cell and Tissue Research* **224**, 269-290.

Technau, U., Miller, M. A., Bridge, D. & Steele, R. E. 2003 Arrested apoptosis of nurse cells during Hydra oogenesis and embryogenesis. *Developmental Biology* **260**, 191-206.

Technau, U., Rudd, S., Maxwell, P., Paul, M. K., Saina, M., Grasso, L. C., Hayward, D. C., Sensen, C. W., Saint, R., Holstein, T. W. & Miller, D. J. 2005 Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians Trends in Genetics. *Trend in Genetics* **21**.

Thompson, J. & Banaszak, L. 2002 Lipid-Protein interactions in lipovitellin. *Biochemistry* **41**, 9398-9409.

Thompson, R. C. & Ohlsson, K. 1986 Isolation, properties, and complete amino acid sequence of Human secretory leukocyte proteinase inhibitor of leukocyte elastase. *Proccedings of the National Academy of Sciences USA*. **83**, 6692-6696.

Tibballs, J. 2006a Australian venomous jellyfish, envenomation syndromes, toxins and therapy. *Toxicon* **48**, 830-859.

Tibballs, J., Williams, D. & Sutherland, S. K. 1998 The effects of antivenom and verapamil on the haemodynamic actions of *Chironex fleckeri* (box jellyfish) venom. *Anaesthesia and Intensive Care* **26**, 40-45.

Torres, M., Aguilar, M. B., Falcon, A., Sanchez, L., Radwan, F. F. Y., Burnett, J. W., Heimer-de la Cotera, E. P. & Arellano, R. O. 2001 Electrophysiological and hemolytic activity elicited by the venom of the jellyfish *Cassiopea xamachana*. *Toxicon* **39**, 1297-1307.

Turk, T. 1991 Cytolytic Toxins from Sea anemones. *Journal of Toxicology-Toxin Reviews* **10**, 223-262.

Underwood, A. H. & Seymour, J. E. 2007 Venom ontogeny, diet and morphology in *Carukia barnesi*, a species of Australian box jellyfish that causes Irukandji syndrome. *Toxicon* **49**, 1073-1082.

Van Loon, L. C. & Van Strien, E. A. 1999 The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physioogical and Molecular Plant Pathology*. **55**, 85-97.

Wagstaff, S. C. & Harrison, R. A. 2006 Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel α9β1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins,renin-like aspartic proteases. *Gene* **377**, 21-32.

Walker, M. J. A., Martinez, T. T. & Godin, D. V. 1977a Investigations into Cardiotoxicity of a Toxin from Nematocysts of Jellyfish, *Cyanea capillata*. *Toxicon* **15**, 339-346.

Wang, W., Omori, M., Hayashibara, T., Shimoike, K., Hatta, M., Sugiyama, T. & T., F. 1995 Isolation and characterization of a mini-collagen gene encoding a nematocyst capsule protein from a reef-building coral, *Acropora donei*. *Gene* **23**, 195.

Watson, G. M. & Mirethibodeaux, P. 1994 The Cell Biology of Nematocysts. *International Review of Cytology - a Survey of Cell Biology, Vol 156* **156**, 275-300.

Weber, K. & Osborn, M. 1969 The Reliability of Molecular Weight Determinations by Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. *Journal of Biological Chemistry* **244**, 4406-4412.

Wegrzyn, J. L., Drudge, T. M., Valafar, F. & Hook, V. 2008 Bioinformatic analyses of mammalian 5'- UTR sequence properties of mRNAs predicts alternative translation initiation sites. *BMC Bioinformatics* **9**, doi:10.1186/1471-2105-9-232.

Welton, R. E. & Burnell, J. N. 2005 Full length nucleotide sequence of a Factor V-like subunit of oscutarin from *Oxyuranus scutellatus scutellatus* (coastal Taipan). *Toxicon* **46**, 328-336.

Werner, B. 1975 Structure and Life-History of Polyp of *Tripedalia cystophora* (Cubozoa, Class-Nov Carybdeidae) and Its Importance for Evolution of Cnidaria. *Helgolander Wissenschaftliche Meeresuntersuchungen* **27**, 461-504.

Whitfield, C. W., Band, M. R., Bonaldo, M. F., Kumar, C. G., Liu, L., Pardinas, J. R., Robertson, H. M., Soares, M. B. & Robinson, G. E. 2002 Annotated Expressed Sequence Tags and cDNA Microarrays for Studies of Brain and Behavior in the Honey Bee. *Genome Research* **12**, 555-566.

Wiedow, O., Schroder, J. M., Gregory, H., Young, J. A. & Christophers, E. 1990 Elafin: an elastase-specific inhibitor of human skin: Purification, characterization, and complte aminoacid sequence. *Journal of Biological Chemistry* **265**, 14791-14795.

Wilkins, A. S. 2000 Hydra: an out-group moves towards the center. *BioEssays* **22**, 200-201.

Williamson, J. 1992 Current Challenges in Marine Envenomation - an Overview. *Journal of Wilderness Medicine* **3**, 422-431.

Wilson, D. a. P., R and Zhou, Y and Talbot, C and Vogel, C and Madera, M and Chothia, C and Gough, J. 2009 SUPERFAMILY- Sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research* **37**, 380-386.

Wiltshire, C. J., Sutherland, S. K., Fenner, P. J. & Young, A. R. 2000 Optimization and preliminary characterization of venom isolated from 3 medically important jellyfish: the box (*Chironex fleckeri)*, Irukandji (*Carukia barnesi*),

and blubber (*Catostylus mosaicus*) jellyfish. *Wilderness & Environmental Medicine* **11**, 241-250.

Winkel, K. D., Hawdon, G. M., Fenner, P. J., Gershwin, L. A., Collins, A. G. & Tibballs, J. 2003 Jellyfish antivenoms: Past, present, and future. *Journal of Toxicology-Toxin Reviews* **22**, 115-127.

Winkel, K. D., Tibballs, J., Molenaar, P., Lambert, G., Coles, P., Ross-Smith, M., Wiltshire, C., Fenner, P. J., Gershwin, L. A., Hawdon, G. M., Wright, C. E. & Angus, J. A. 2005 Cardiovascular actions of the venom from the Irukandji (*Carukia barnesi*) jellyfish: Effects in human, rat and guinea-pig tissues in vitro and in pigs in vivo. *Clinical and Experimental Pharmacology and Physiology* **32**, 777-788.

Winter, K. L., Isbister, G. K., Schneider, J. J., Konstantakopoulos, N., Seymour, J. E. & Hodgson, W. C. 2008 An examination of the cardiovascular effects of an 'Irukandji' jellyfish, *Alatina nr mordens*. *Toxicology Letters* **179**, 118-123.

Winter, K. L., Isbister, G. K., Seymour, J. E. & Hodgson, W. C. 2007 An in vivo examination of the stability of venom from the Australian box jellyfish *Chironex fleckeri*. *Toxicon* **49**, 804-809.

Wu, i., Skolnick, J. & Zhang, Y. 2007 Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biology* **5**.

Xiao, T., DeCamp, D. L. & Spran, S. R. 2000 Structure of a rat alpha 1-macroglobulin receptor-binding domain dimer. *Protein Sciences* **9**, 1889-1897.

Yan, L., Fei, K. Y., Zhang, J. S., Dexter, S. & Sarras, M. P. 2000 Identification and characterization of hydra metalloproteinase 2 (HMP2): a meprin-like astacin metalloproteinase that functions in pedalia morphogenesis. *Development* **127**, 129-141.

Yang, Y. Z., Cun, S. J., Peng, L. S., Xie, X. J., Wei, H. W., Yang, W. L. & Xu, A. L. 2003b cDNA cloning, identification and characterization of a novel cystatin from the tentacle of *Cyanea capillata*. *Biochimie* **85**, 1033-1039.

Yang, Y. Z., Cun, S. J., Xie, X. J., Lin, J. H., Wei, J. W., Yang, W. L., Mou, C. Y., Yu, C. L., Ye, L. T., Lu, Y., Fu, Z. Y. & Xu, A. L. 2003a EST analysis of gene expression in the tentacle of *Cyanea capillata*. *FEBS Letters* **538**, 183-191.

Yang, Z. 2008 I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**.

Yu, H. H., Xing, R. G., Liu, S., Li, C. P., Guo, Z. Y. & Li, P. C. 2007 Studies on the hemolytic activity of tentacle extracts of jellyfish *Rhopilema esculentum Kishinouye*: Application of orthogonal test. *International Journal of Biological Macromolecules* **40**, 276-280.

Zhang, M. L., Fishman, Y., Sher, D. & Zlotkin, E. 2003 Hydralysin, a novel animal group-selective paralytic and cytolytic protein from a noncnidocystic origin in hydra. *Biochemistry* **42**, 8939-8944.

Zhu, S. & Tytgat, J. 2004 The scorpine family of defensins: gene structure, alternative polyadenylation and fold recognition. *Cellular and Molecular Life Sciences* **61**, 1751-1763.