Characterisation of Metalloproteinases from Myanmar Russell's Viper

Thesis submitted in accordance with the requirements of the Chulalongkorn University and University of Liverpool for the degree of Doctor in Philosophy in Biomedical Sciences and Biotechnology

By

Khin Than Yee

June 2017

Acknowledgements

I would like to take this opportunity to express my appreciation and gratitude to program managers and staffs of the Joint Program in Biomedical Sciences and Biotechnology, in particular to the committee members of the scholarship for International Students from Faculty of Medicine, Chulalongkorn University for their generous funding.

I would like to thank my supervisors, Prof. Ponlapat Rojnuckarin from Chulalongkorn University and Dr. Mark Wilkinson and Dr. Olga Vasieva from University of Liverpool for their kind support and guidance in getting me through the duration of my PhD.

None of this research would have been possible without the generous permission to allow me to study in Thailand and United Kingdom by Dr. Kyaw Zin Thant, Director General, Department of Medical Research, Ministry of Health and Sports, Myanmar.

I should not to fail to be grateful to all officers and staffs from the Snake Farm, Myanmar Pharmaceutical Factory, Yangon, Myanmar and Dr. Lawan Chanhome, Head of the Snake Farm, Queen Saovabha Memorial Institute, The Thai Red Cross Society, Thailand for their enthusiastic help in taking samples of venom and venom glands in Myanmar.

Thank you to all of my friends who have given me a sympathetic ear, moral support and encouragement throughout the past four years of my research.

Special thanks to my family members whose listening, understanding has proven invaluable to have full attention over my PhD study.

I declare that this thesis entitled:

Characterisation of Metalloproteinases from Myanmar Russell's Viper

Is entirely my own work

Candidate: Khin Than Yee

Supervisors: Prof. Ponlapat Rojnuckarin (Chulalongkorn University)

Dr. Mark Wilkinson (University of Liverpool)

Dr. Olga Vasieva (University of Liverpool)

Abstract

Russell's viper bites are a major public health problem in tropical and subtropical regions. In Myanmar, a Russell's viper (*Daboia russellii siamensis*) bite has a 60% morbidity rate and 8.2% fatality rate. Most victims encounter severe bleeding, renal failure and capillary leakage and the bite can possibly lead to death. Snake venom metalloproteinases (SVMPs) are the major components of the Viperidae venom and all mentioned lethal effects of the bites are attributed to these. The only available and partially effective agent for the treatment of the toxic effects is antivenom. Antivenom therapy is not always effective towards small toxins, however, and it can also provoke an anaphylactic response. The development of new therapeutic approaches is becoming increasingly important therefore.

For the analysis of SVMP transcripts from Myanmar Russell's viper, Next-Generation Sequencing (NGS) of mRNA from venom glands derived from 2 male snakes and 1 female snake was performed on an Illumina HiSeq2000 platform. *De novo* assembly of the reads was performed using Trinity software and the transcripts were annotated through Blastn against the collection of NCBI nucleotide sequences defined by the key-words ('venom' and 'serpents') search. Blastx hit results against the UniProtKB/Swiss-Prot (swissprot) database were also used for annotation of the transcripts. The abundance distribution (in term of FPKM value) of SVMPs toxin transcripts: disintegrin (75%), P-III SVMPs (25%) and P-II SVMPs (0.002 %), were the same for both male and female samples. No P-I SVMP transcripts were detected in the present analysis. A comparison of the contents of SVMP transcripts in adult male and female venom glands showed some gender-related differences. For example, a disintegrin transcript isoform (Dis 1b) was highly expressed only in the female venom gland, and some P-III SVMP isoforms (P-III 6, 7a, 7b) were only expressed at low level in the male venom glands. The P-II SVMP transcripts were expressed as different isoforms in male and female animals, which could reflect a sex-dimorphism of viper venom biological activities. This finding would support a requirement to use combined venoms of both sexes for preparation of antivenom. In addition to SVMP transcripts, mRNAs of novel tripeptide SVMP inhibitors (SVMPIs) were also

discovered. These endogenous inhibitors have potential as a new treatment modality for neutralization of the effect of SVMP toxins.

Two major snake SVMPs, RVV-X and daborhagin, were purified from Myanmar Russell's viper venom using a new purification strategy. Moreover, the two novel endogenous tripeptides identified in transcript analysis, pERW and pEKW were identified and isolated from the crude venom. Both purified SVMPs showed caseinolytic activity. Additionally, RVV-X displayed specific proteolytic activity towards gelatin and Daborhagin showed potent fibrinogenolytic activity. These activities were inhibited by metal chelators. Notably, synthetic versions of the peptide inhibitors, pERW and pEKW, completely inhibited the gelatinolytic and fibrinogenolytic activities of the respective SVMPs when used at 5 mM concentration (estimated molar ratio of SVMP to tripeptide was 1:500). These complete inhibitory effects suggest that these tripeptides deserve further study as candidates for new therapeutic treatment against Russell's viper envenomation.

Contents

List of Figures

List of Tables

List of abbreviations

5'RACE	5' Rapid Amplification of cDNA Ends
AAP	Abridged Anchor Primer
ACLH	*Agkistrondon contortrix laticinctus* haemorrhagic toxin
ADAM	A Disintegrin And Metalloproteinase
ADAMDEC-1	Decysin-1
ADAMTS	ADAM with thrombospondin type-1 motif
ANP	Atrial Natriuretic Peptide
BJ46a	Anti-haemorrhagic factor from serum of *Bothrops jararaca*
BPP	Bradykinin-potentiating peptide
cDNA	Complementary deoxyribonucleic acid
CDS	Coding Sequence
cfu	Colony-forming unit
CLPs	C-type lectin-like proteins
cMSF	Chinese mamushi serum factor
CNP	C-type Natriuretic Peptide
CRD	Carbohydrate recognition domain
CRISP	Cysteine-rich secretory protein
CVO	*Crotalus viridis oreganus*
DbTx-B	Daboiatoxin-B
dCTP	Deoxycytidine triphosphate
DEPC	Diethyl pyrocarbonate
DM40, DM43	Antibothropic complex from *Didelphis marsupialis* serum
dNTP	deoxynucleotides
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetra acetic Acid
EGF	Epidermal growth factor
EGTA	Ethylene glycol-bis (2-aminoethylether)-*N,N,N',N'*-tetraacetic acid
ESI	Electrospray Ionisation
ESTs	Expressed sequence tags
FPKM	Fragments Per Kilobase of exon per Million fragments mapped
GDP	Gross domestic product

GFC	Gel Filtration Chromatography
gi	GenInfo Identifier
GSP	Gene-specific primer
H2, H3, H4, H6	Haemorrhagic SVMPs from mamushi venom
HIC	Hydrophobic Interaction Chromatography
HR1A, HR1B	Haemorrhagic factors from habu venom
HSF	Habu serum factor
HUVECs	Human umbilical vein endothelial cells
jMSF	Japanese mamushi serum factor
LC	Liquid Chromatography
LINEs	Long Interspersed Elements
m/z	Mass to charge ratio
M12B	Peptidase subfamily M12B (adamalysin)
MALDI	Matrix Assisted Laser Desorption Ionisation
MBP	Mannose-binding protein
MDC	Metalloproteinase, disintegrin-like, cysteine-rich domains
MMPs	Matrix metalloproteinases
mRNA	Messenger RNA
MRV	Myanmar Russell's Viper
MS	Mass Spectrometry
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
ORF	Open reading frame
PCR	Polymerase chain reaction
pEKW	Pyroglutamate-lysine-tryptophan
pENW	Pyroglutamate-asparagine-tryptophan
pEQW	Pyroglutamate-glutamine-tryptophan
pERW	Pyroglutamate-arginine-tryptophan
PLA_2	Phospholipase A_2
RP-HPLC	Reversed phase high performance liquid chromatography
RSEM	RNA-Seq by Expectation Maximization
RT	Reverse Transcriptase

RV-4	Phospholipase A_2 from Taiwan Russell's viper
RVBCMP	Russell's viper basic coagulant metalloproteinase
RVV-X	Russell's Viper Venom factor X activator
RVV-X-HC	RVV-X heavy chain
RVV-X-LC1	RVV-X light chain 1
RVV-X-LC2	RVV-X light chain 2
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SINEs	Short Interspersed Elements
Snaclecs	Snake C-type lectins
SOC	Super Optimal broth with Catabolite repression
SPs	Serine Proteases
SV-LAAOs	Snake venom L-amino acid oxidases
SVMPI	Snake Venom Metalloproteinase Inhibitor
SVMPs	Snake Venom Metalloproteinases
TCA	Trichloroacetic acid
TdT	Terminal deoxynucleotidyl Transferase
TE	Transposable element
TFA	Trifluoroacetic acid
TOF	Time-of-Flight
TSR	Thrombospondin type-1 Repeat
VLAIP	*Vipera lebetina* apoptosis-inducing protease
VRH-1	A haemorrhagin from *Vipera russelli russelli* venom
VRR-73	*Vipera russelli russelli* -73
WHO	World Health Organization

Chapter 1. Introduction

Snakebites are a major public health problem in tropical and subtropical regions. The World Health Organization has recognised the snakebite as a Neglected Tropical Disease. A lack of reliable incidence data owing to unreported cases has resulted in the disease being neglected. In addition, snake antivenoms have become scarce or non-existent as poor economic viability has forced some manufacturers to leave the market, and others to downscale production (1). Thus, it is an important issue for those tropical countries with the lowest per capita government expenditure on health and the lowest per capita GDP that suffer the highest snakebite mortality. It is estimated that at least 8,832 cases of envenoming and over 30 deaths each year in Myanmar result from snakebites (2).

Severe cases of snake bite envenoming are inflicted by species of the family Elapidae (cobras, kraits, mambas and sea snakes) or the family Viperidae (rattlesnakes, lance-headed pit vipers, and true vipers). Russell's viper, a true viper, causes high morbidity and mortality across the Asian countries including parts of mainland China, western Cambodia, Thailand, Myanmar, Bangladesh, southern Indonesia, Taiwan, Pakistan, India and Sri Lanka (3). The Russell's viper gives bleeding manifestations which lead to the death of the victims. The predominant coagulopathic effects of viper bites are caused by the major constituents of the venom: proteinases. Investigations on snake venom composition started a long time ago. However, recent developments in the techniques and strategies for analysis of snake venom proteomes and venom gland transcriptomes has led to a better understanding of the important physiological processes in which venom components are involved. A thorough assessment of the toxin composition of snake venom (venomics) directly (through proteomics-centered approaches) or indirectly (via high-throughput venom transcriptomic and bioinformatic analysis) is relatively rapid and cost-effective (4). In addition, both approaches are feasible to explore in depth the molecular diversity of venoms and to analyse various functional aspects of the toxins such as post-translational modifications, proteolytic processing and toxin-target interactions, in addition to the toxin protein identification.

Although the proteomic and transcriptomic data for most vipers are available, the data for some Russell's vipers are still unreported, especially for Myanmar species. This work aims to identify, quantify and characterise the medically important toxin group, the snake venom metalloproteinases, from Myanmar Russell's viper at transcript and protein levels for a better understanding of their domain structural alteration which leads to a diverse array of functions for envenomation. Such understanding may suggest better snakebite treatment modalities in the future.

1.1. Literature review

1.1.1. Epidemiology

Russell's viper (*Daboia russellii*) is a medically important snake, variants of which are distributed throughout East and Southeast Asia. There are 5 subspecies of Russell's viper according to minor differences in colouration and markings (see Figures 1.1 and 1.2): *D. r. russelli* in India, *D. r. pulchella* in Sri Lanka, *D. r. siamensis* in Myanmar, Thailand and China, *D. r. formosensis* in Taiwan and *D. r. limitis* in Indonesia (5). In Myanmar, snakebite cases occur in almost all areas of the country but the Mandalay, Pegu, Sagaing, Irrawaddy and Yangon Divisions are mostly affected (6). A Russell's viper bite has a 60% morbidity rate and the fatality rate is 8.2 % in Myanmar (7).

1.1.2. Clinical manifestations of Russell's viper bites

Patients may have symptoms such as moderately severe pain at the site of bite and regional lymph node enlargement, vomiting, early transient syncope, abdominal pain, and drowsiness. The local signs are blistering, necrosis and bleeding from the fang marks. Systemic envenomation is manifested by incoagulable blood, thrombocytopenia, spontaneous systemic bleeding, hypotension, increased capillary permeability and oligouria (8). The causes of death include shock, pituitary and intracranial haemorrhage, massive gastrointestinal haemorrhage and acute renal failure resulting from systemic envenomation. The clinical manifestations vary between the geographical areas, reflecting differences in venom composition within the 5 subspecies (5). The variations in clinical features resulting from Russell's viper bites among the subspecies are listed in Table 1.1.

Figure 1.1. Eastern Russell's vipers (*Daboia siamensis*) (Copyright DA Warrell). (a) Specimen from Myanmar; (b) Specimen from Thailand; (c) Specimen from East Java, Indonesia; (d) Specimen from Flores, Indonesia. Adapted from Guidelines for the management of snake-bites, David A Warrell, WHO 2010.

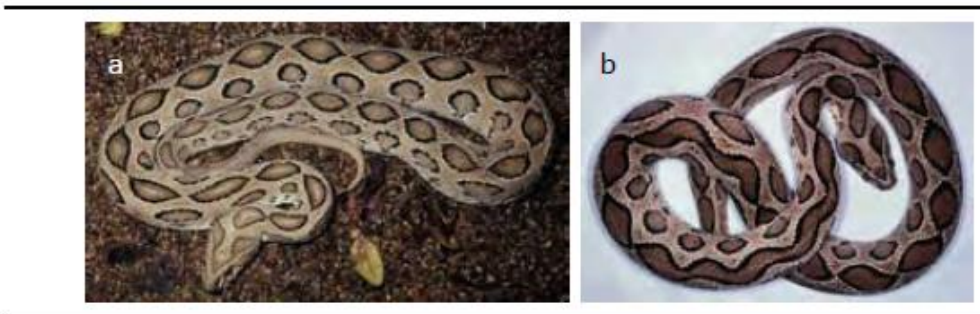


Figure 1.2. Western Russell's vipers (*Daboia russelii*) (Copyright DA Warrell). (a) Specimen from southern India; (b) Specimen from Sri Lanka. Adapted from Guidelines for the management of snake-bites, David A Warrell, WHO 2010.

Table 1.1. Variations in clinical manifestations of Russell's viper bites in various countries. Data from Belt et al, 1997 (3); Hung et al, 2002 (9); Antonypillai et al, 2011 (10); Kularatne, 2003 (11); Kularatne et at, 2014 (12) studies are added to the data summarized in Warrell (1989) (13).

Symptom	Myanmar *D.r.siamensis* (n=631)	Thailand *D.r.siamensis* (n=82)	Sri Lanka *D.r.pulchella* (n=450)	India *D.r.russelii* (n=1608)	Taiwan *D.r.formosensis* (n=18)	Indonesia *D.r.limitis* (n=26)
Coagulopathy	++	++	+	++	++	++
Renal failure	++	+	++	+	+	+
Pituitary infarction	++	-	+	+	-	-
Haemolysis	-	+	++	+	+	-
Neuromyotoxicity	-	-	++	+	-	+
Oedema	++	-	-	-	+	+
Shock	++	-	+	+	-	-
Thrombosis	-	-	-	-	+	-

n = total number of patients

From Table 1.1, haemostatic abnormalities are seen in all countries but coagulopathy was less marked in Sri Lanka envenoming than in any other country. Renal failure was also a common clinical manifestation in all countries but most significant in Myanmar and Sri Lanka. Pituitary haemorrhage has so far been reported in Myanmar, Sri Lanka and India, but is most severe in Myanmar. Intravascular haemolysis was most prominent in Sri Lanka but has been found in Thailand, India and Taiwan. Neuromyotoxicity was a prominent clinical feature in Sri Lanka and also seen in India and Indonesia. Edema and primary shock the most well-known manifestations in Myanmar. However, arterial thrombosis is most commonly described in Taiwan.

Such variation in clinical manifestations in envenoming in different countries indicates the geographical variation in venom composition, as well as suggesting that there are differences in the sequences and quantities of the active components in the venoms. There may also be variations in responses of people who are bitten, or perhaps variations in the clinical reporting. The immune response towards snake venoms depends on the physiochemical characteristics of individual toxins, such as structure and molecular masse, on their relative abundance in the venom, and the dose of venom that entered into the body. Also, host factors, such as the capacity of the immune system to recognize each toxin, individual genetic background and the

ability of process the toxins also need to considered in the clinical outcomes following snake envenomation (14).

1.1.3. Venom composition

Russell's viper venom contains many different components: toxic and non-toxic, enzymatic and non-enzymatic. Approximately 90% of the dry weight of venom is protein (5). Other components in the venom are small peptides, metallic cations, carbohydrates, nucleoside, biogenic amines, free amino acids and lipids (15, 16). A proteomic study of Siamese Russell's viper venom from Myanmar revealed a wide diversity of several toxins belonging to at least 6 protein families: serine proteinases, metalloproteinases, phospholipase A_2 (PLA_2), L-amino acid oxidases (LAAOs), vascular endothelial growth factors (VEGFs) and C-type lectin-like proteins (CLPs) (17). This proteomic study did not describe the percentage distribution of the toxins but only the number of different proteins or isoforms were identified. Eleven isoforms of blood coagulation factor V activator-like serine proteases (RVV-V) and 2 homologs of halystase from the serine protease family; 6 isoforms of RVV-X from the metalloproteinase family; 10 single chains, one dimeric and 11 complexed with other proteins from the PLA_2 family; 6 isoforms of LAAO; 2 isoforms of VEGF; and 2 proteins of CLPs were detected from MRV proteome. However, recent proteomic studies on Russell's viper venom from Indian (18, 19) and Sri Lankan (20) snakes elucidated the relative abundance of toxin families (Table 1.2). There are 55 proteins belonging to 13 distinct snake venom families; 63 different proteins belonging to 12 families; and 41 different venom proteins belonging to 11 different protein families in Western India (Mumbai) RVV venom, Southern Indian (Tamil Nadu) RVV venom, and Sri Lankan RVV venom, respectively. The proteomes of Russell's viper snake venoms from Thailand and Taiwan have not been studied in detail, particularly about the composition and identification of toxins contained, but 2D gel images of those venoms showed similar protein patterns in *V. russelli formosensis* and *V. russelli siamensis* (21). The proteomic studies on other viperid snakes exhibit intragenus variations in different toxin protein families: 11 toxin families in *Bitis gabonica*, 9 in *Bitis nasicornics*, 8 in *Bitis caudalis* (22), 9 in *V. ammodytes meridionalis* and 8 in *V. ammodytes ammodytes* respectively (23). In addition, a comparative

characterisation of Russell's viper venoms from the Indian peninsula showed significant regional variation in the constituents of venoms, and the eastern regional venom is the most lethal of all the venoms of the Indian peninsula (24)**.** Variations in venom composition are attributed to multiple levels of regulation on transcription, translation, and posttranslational modification of toxins as an adaptation to different kinds of prey in different geographical locations. This variation results in significant differences in venom-induced pathology and lethality between related snake species and can undermine the efficacy of antivenom therapy (25).

Differences in venom composition of Russell's viper from India and Sri Lanka was shown by their complete venomic studies, while the venomics of Myanmar, Taiwan and Thailand species have only been partially studied. However, the difference in venom composition between Myanmar and Thailand species has not been elucidated and can only be estimated indirectly from the enzymatic assay of those crude venoms. The crude venom of Myanmar Russell's viper had high protease activity (26), whereas Thailand Russell's viper crude venom was high in phospholipase A_2 activity (27).

Table 1.2. Relative abundance of toxin families from different Russell's vipers determined by proteomic studies

Toxin family	Western India *D. russelii*	Southern India *D. r. russelii*	Sri Lanka *D. russelii*
Phospholipase A_2 (PLA_2)	32.5%	23.8%	35.0%
Metalloproteinases (SVMPs)	24.8%	9.5%	6.9%
Serine proteases (SVSPs)	8.0%	17.5%	16.0%
Cysteine-rich secretory proteins (CRISPs)	6.8%	11.0%	2.0%
Disintegrin	4.9%	3.2%	-
Venom nerve growth factors (VNGFs)	4.8%	1.6%	3.5%
C-type lectin-like proteins (CLPs/Snaclec)	1.8%	7.5%	22.4%
Vascular endothelial growth factors (VEGFs)	1.8%	3.2%	-
Phosphodiesterases (PDEs)	1.4%	3.2%	1.3%
Nucleotidases (NTs)	0.4%	4.8%	3.0%
L-amino acid oxidases (LAAOs)	0.3%	7.9%	5.2%
Phospholipase B (PLB)	0.1%	-	0.1%
Kunitz-type serine protease inhibitor (KSPI)	12.5%	4.8%	4.6%

1.1.4. Venom Biochemistry

The toxins of snake venom affect the haemostatic, cardiovascular, nervous systems following envenoming and cause tissue necrosis. Metalloproteinases, serine proteases, C-type lectins, disintegrins, and phospholipases affect haemostasis by activating or inhibiting coagulant factors or platelets, and disrupting vascular endothelium. Metalloproteinases damage the vascular endothelium and cause spontaneous systemic bleeding. Procoagulant enzymes such as thrombin-like fibrinogenases, activators of prothrombin, factor V, X, and XII, and endogenous plasminogen from venom disturb the coagulation pathway. Anticoagulant venom phospholipases A_2 hydrolyse or bind to procoagulant phospholipids and inhibit the prothrombinase complex. L-amino acid oxidases activate platelets via H_2O_2 production (28). Thus, severe bleeding results from the combination of consumption coagulopathy, anticoagulant activity, decreased and/or dysfunctional platelets, and vessel wall damage. The biological activities of known venom proteins from Russell's vipers are listed in Table 1.3. Many major venom proteins from different Russell's vipers were purified and their biological activities were characterized. However, some proteins such as disintegrin, CRISP and phospholipase B from Russell's viper venom have not yet been studied thoroughly.

Table 1.3. Known venom proteins from Russell's vipers with their biological activities

Toxin family	Individual toxin	Functions
SVMPs	RVV-X (29, 30)	Activation of Factor X, renal failure through intravascular clotting in renal microcirculation
	VRR-73 (31)	Activation of plasminogen, fibrinolysis, haemorrhage
	VRH-1 (32)	Severe lung haemorrhage
	RVBCMP (33)	Distinct liver haemorrhage, fibrinogenolytic and procoagulant activities
	Daborhagin-K Daborrhagin-M (34)	Severe haemorrhage Daborrhagin-M has hydrolytic effect on fibrinogen, fibronectin, type IV collagen
	MRV-75 (26)	Haemorrhage, edema, myonecrotic, coagulant, finbrin(ogen)olytic activities

Table 1.3. Known venom proteins from Russell's vipers with their biological activities (Cont.)

Toxin family	Individual toxin	Functions
SVSPs	RVV-V (35)	Activation of Factor V
	Russelobin (36)	Thrombin-like proteinase causing defibrination
	RV-FVP (37)	Fibrinogenolytic, plasma clotting activities
PLA_2	Daboiatoxin (DbTx) (38)	Neurotoxic, myotoxic, cytotoxic, edema-inducing, indirect-haemolytic activities
	VRV-PL-IIIb (39)	inhibition of platelet aggregation, edema
	VRV-PL-V (40)	Neurotoxic, edema, myonecrosis, anticoagulant Haemorrhage in lung & liver
	VRV-PL-VI (41)	Neurotoxic, edema, anticoagulant, haemorrhage in lung & liver
	VRV-PL-VIIIa (42)	Neurotoxic, myonecrosis, edema and anticoagulant, haemorrhage in lung & liver
	Viperotoxin F (RV-4 and RV-7) (43)	Neurotoxic
	RVV-7 (44)	Cytotoxic, inhibit platelet aggregation
	Daboxin P (45)	Anticoagulant activity by targeting Factor X and Xa
KSPI	Rusvikunin complex (46)	Anticoagulant activity
	DrKIn-II (47)	Antifibrinolytic activity (slow and tight-binding plasmin inhibitor)
CRISP	-	Smooth muscle inhibition (48)
Disintegrin	-	Platelet aggregation inhibitors (49)
NGFs	NGF from *V. r. russelli* (50)	Not known
Snacle	RV snaclec (51)	Anticoagulant activity
	Daboialectin (52)	Cytoskeletal damage and apoptosis in human lung cancer cells
VEGF	VEGF165 (53)	Proliferation of vascular endothelial cells, hypotension on rat arterial blood pressure
PDEs	PDE from Myanmar Russell's viper (54)	Edema forming, myonecrotic, haemorrhagic activities
	DR-PDE (55)	Inhibition of platelet aggregation
NTs	Phosphomonodiesterase from MRV (56)	Edema, myonecrotic activity
LAAOs	DrLAO (57)	Inhibit platelet aggregation
PLB	-	Haemolysis, cytotoxic (58)

1.1.5. The venom transcriptome

Transcriptomic profiling of venom gland tissue is an excellent approach to understand the composition and complexity of the snake venom. The transcriptomic study of venom glands could help to achieve the following goals: 1) documentation of toxin profiles in venoms, 2) identification of novel toxins, 3) elucidation of toxin variations, 4) development of antivenoms, 5) understanding of evolution of snakes and their venoms (59). Toxin profiling from the transcriptome is useful for both targeted searches for individual toxins or isoforms within one toxin family and discovery of all toxins attributed to a species.

There are advantages and disadvantages of transcriptomic approaches. The venom gland transcriptome can help to determine not only the expression profile and the evolution of the venom proteins, but also can reveal the lowly-expressed proteins that may not be detected by proteomic approaches, but nevertheless may be important in pathology and which are have pharmaceutical potential (60). Also, variations in venom composition and relative expression of venom proteins can be elucidated by transcriptomics approaches. Moreover, venom complexity can be studied for mechanisms of venom protein evolution and neo-functionalisation. In addition, identification of toxin gene structures has supported the understanding of potential toxin functions and in predicting structure-functional relationships.

However, transcriptomic analysis cannot provide information of post-translational processing of venom toxins. Post-translational modifications such as glycosylation, proteolytic cleavage, oligomerisation may also contribute to important biological activities of toxins. Moreover, transcriptomics cannot indicate exactly which mRNA transcripts are actually translated into proteins (59).

The limitations on transcriptomic are the requirement of suitable and efficient bioinformatic tools to analyse the very large datasets and provide into meaningful interpretations. Further, functional annotation of genes in non-model organisms, such as venomous animals, is based mostly on the annotated genes of model organisms such as yeast, flies, nematodes and mice. These classic model organisms do not have toxin/venomous function and many toxin genes were recruited from

genes of non-venomous function. Thus, it might lead to annotation with incorrect function and this may pose difficulties in differentiating the venomous and non-venomous genes by manual annotation (61).

The first reported venom gland cDNA library obtained using cloning technology was for the viperid *Bothrops insularis*. The resultant cDNA library showed that (a) toxin genes contributed a large percentage of expressed genes in the venom gland, (b) the diversity of toxin genes was higher than that of non-toxin genes, (c) the presence of novel toxin genes and families (62). The cDNA library from Thailand Russell's viper venom glands was constructed using ZAP express cDNA Synthesis Kit and ZAP express cDNA Gigapack III Gold Cloning Kit (Stratagene). This approach showed that phospholipase A_2 were the most abundant transcripts followed by BPP-CNP precursor, serine protease, metalloproteinases and nucleotidase toxin transcripts. There is no other transcriptomic study for other Russell's viper species available (63). The first NGS approach for cataloguing toxin genes from the eastern diamondback rattlesnake (*C. adamanteus*) used 454 pyrosequencing technology. This study further exhibited the ability to detect toxins in very low abundance in the mRNA pool which is more difficult to do using cloning techniques (64). Toxin cataloguing using NGS has an advantage over cloning technology in saving the time and effort needed to generate data with greater coverage. Currently, there are seven snake families from which snake venom transcriptomes have been constructed using both techniques. However, that of Daboia species has not been thoroughly explored (59).

1.1.6. Venom evolution

Snake venoms evolve via a process by which a gene encoding for a normal protein is duplicated and a copy is selectively expressed in the venom gland. The duplication of genes and their subsequent functional divergence, leading to the formation of families of evolutionarily related but functionally distinct genes, is a fundamental process of adaptive evolution (65). Large venom-specific multigene families have been identified including PLA_2, serine proteases, Kunitz-type serine protease inhibitors and metalloproteinases (66). PLA_2 toxicity is mediated through pharmacological sites on the surface of the enzyme. These pharmacological sites, which are distinct from the catalytic site of the enzyme, bind to the specific target

proteins. These regions under positive selection on the protein surface are responsible for generating toxic functions through a process of neo-functionalisation (67). Mutations and exon rearrangements leading to an encoded protein domain loss also plays a key role in the neo-functionalisation of duplicated genes.

The snake venom metalloproteinase family comprises a large group of related toxins generated through domain loss (68) resulting from the appearance of a stop codon in a precursor-gene. P-III class of SVMPs have a modular structure containing metalloproteinase (M), disintegrin-like (D) and cysteine-rich (C) domains. The structure of *Echis ocellatus* pre-pro EOC00089-like PIII-SVMP gene comprises 12 exons interrupted by 11 introns. PII-SVMPs in viperid venoms are minimised into a disintegrin gene by loss of intron 1 in the genes encoding α-subunits of dimeric disintegrins, and in short disintegrin genes. The dimeric disintegrin α-subunit ancestral gene may have derived from a β-subunit precursor by genetic rearrangement resulting in the removal of the region comprising exons 3-10 and introns 2-10, and the fusion of exons 2 and 11 (69). The sites under selection in toxin genes may be influenced by a combination of factors including an animal's feeding habits and their environment; and bio-geographical factors including population expansions, bottlenecks and vicariance. P-III SVMPs are widely distributed in the five snake families of Colubridea, Viperidae, Elapidae, Atractaspididae, and Colubridae, while P-II disintegrins have been only found in venoms of Viperidae. This suggests that distintegrins emerged from after the split of Viperidae and Elipidae, along another lineage, before the separation of the Viperidae subfamilies. The disintegrin subfamilies segregate distinctly in African and Eurasian Viperinae, expressing mainly dimeric and short disintegrins, and in Asian and New World Crotalinae, in whose venoms both large and medium-sized disintegrins are significantly seen (70).

1.1.7. ADAM, ADAMTS and snake venom metalloproteinases

Snake venom metalloproteinases (SVMPs) are phylogenetically most closely related to the mammalian ADAM (a disintegrin and metalloproteinase) family and ADAMTS (ADAM with thrombospondin type-1 motif) family of proteins (71). They belong to the adamalysins/reprolysins or the M12B clan of metalloendopeptidases (72).

An ADAM consists of a prodomain, a metalloproteinase domain (M), a disintegrin-like domain (D), a cysteine-rich domain (C), an EGF-like domain, a short connecting linker, a transmembrane domain (TM), and a cytoplasmic tail (Figure 1.3). There are no EGF domains in ADAM 10 and 17, and the MDC domains are followed by the TM segment. ADAM 9, 12 and 28 have splicing variants and are expressed as soluble proteinase without the transmembrane and cytoplasmic regions. The ADAMDEC-1 (decysin-1), regarded as an ADAM, is comprised of an M domain and a short disintegrin-like domain.

The adamalysins subfamily contains the class P-III SVMPs and the ADAM-TS family. P-III SVMPs have a modular structure homologous to the MDC domains of the membrane-anchored ADAMs. The SVMPs are referred to as snake ADAMs as they have been derived from an ancestral gene encoding closely related to ADAM 7, 28 and ADAMDEC-1. The ADAM-TS members consist of metalloproteinase (M), disintegrin-like (D), thrombospondin type-1 repeat (TSR) motif (T1), cysteine-rich (C) and spacer (S) domains. The D domain of ADAMTSs actually have an ADAM_CR domain (D*) instead of a classic “disintegrin-like” domain.

Matrix metalloproteinases (MMPs), a sub family of ADAM, share topologically similar zinc metalloproteinases domain with ADAMTs and SVMPs (73-75).

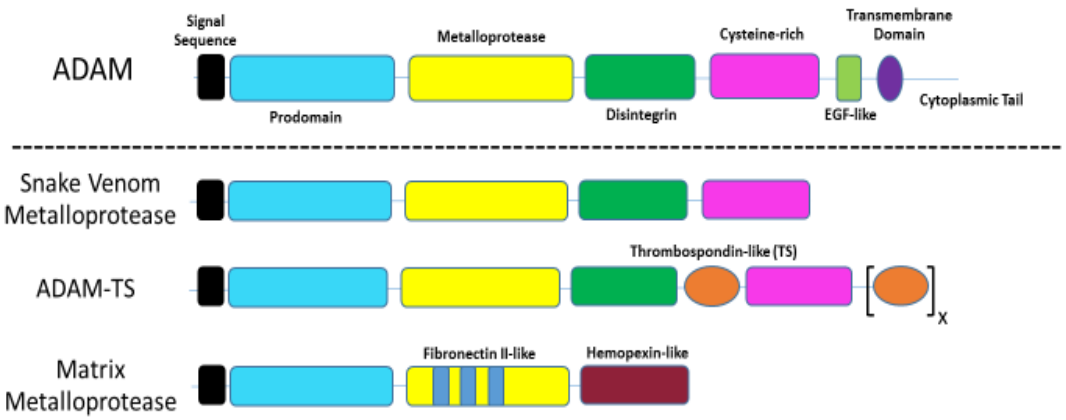


Figure 1.3. The topography of the ADAMs and related metalloproteinases. The domain structures of the ADAMs, SVMPs, ADAM-TS and MMP families. Adapted from Seals & Courtneidge, 2003.

1.2. Snake venom metalloproteinases

Snake venom metalloproteinases (SVMPs) are amongst the key enzymes contributing to local and systemic haemorrhage in viper bites. The SVMPs constitute major components of the venom with the SVMP content varying from 11% to over 65% of the total protein in Viperidae venoms (76).

SVMPs are classified into P-I to P-III classes according to their domain organisation with different molecular weights: Class I (P-I) is the simplest class of enzyme containing only a pro domain and a metalloproteinase domain (20-30 kDa); Class II (P-II) contains a pro domain and a metalloproteinase domain followed by disintegrin domain (30-60 kDa); Class III (P-III) contains a pro, metalloproteinase, disintegrin-like and cysteine-rich domain (60-100 kDa) (77). P-III SVMPs are further divided into subclasses depending on their distinct post-translational modifications, such as homo-dimerization (P-IIIc) or proteolysis between the metalloproteinase and disintegrin-like domains (P-IIIb). The heterotrimetic class of SVMPs (P-IIId), previously called P-IV SVMPs, contains additional snake C-type lectin-like domains. Similarly, the post-translationally modified forms of P-II SVMPs are subclassified into P-IIa, P-IIb and PIIc (68, 78) (Figure 1.4).

Through their proteolytic action on a variety of substrate proteins, the SVMPs show a wide spectrum of biological activities such as haemorrhage, fibrin(ogen)olysis, prothrombin activation, factor X activation, apoptosis, platelet aggregation inhibition, pro-inflammation and inactivation of blood serine proteinase inhibitors (77). This diversity in biological functions is a consequence of the varied structural domains of the SVMPs.

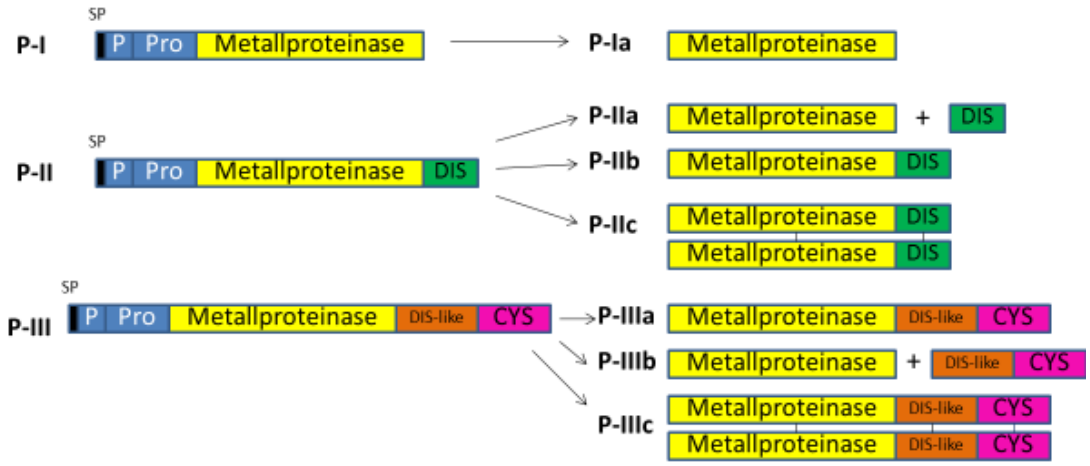


Figure 1.4. Schematic of SVMP classes and their post-translationally modified forms in venom. SP= signal peptide, P= pre-domain, Pro= pro-domain, Metalloproteinase= metalloproteinase domain, DIS= disintegrin domain, DIS-like= disintegrin-like domain, CYS= cysteine-rich domain. Adapted from Casewell *et al* (2011).

1.2.1. The metalloproteinase domain

The catalytic metalloproteinase domains of SVMPs range from 200 to 210 residues in length and contain the consensus sequence HEXGHXXGXXHD (71). The overall shape of the protein is an oblate ellipsoid with a central five-stranded β-sheet mixed with α-helices along with the conserved Met-Turn structure between the αD and αE helices. The structure is composed of an upper and lower domain with the substrate-binding cleft running between them. In the active-site cleft (Figure 1.5), the catalytic zinc ion is co-ordinated with His 142, His 146 and His 152 and to a water molecule providing catalytic activity (79). Structural differences account for the activity disparities between the haemorrhagic and the non-haemorrhagic SVMPs. The ability to induce haemorrhage is related to backbone flexibility in specific surface regions of the metalloproteinases. The haemorrhagic SVMPs exhibit enhanced motions in the loop at residues 156–165, whereas the non-haemorrhagic SVMPs have a dynamic loop at residues 167–175 (80). In addition to the protein secondary structure, post-translation modifications such as glycosylation have an important role in the haemorrhagic activity of SVMPs. The amino acid sequence alignment of the catalytic

domain of jararahagin and ACLH indicates a single potential glycosylation site (NCYS) at the C-terminus (81, 82). The haemorrhagic property of these haemorrhagic SVMPs on mouse skin is lost after deglycosylation (83).

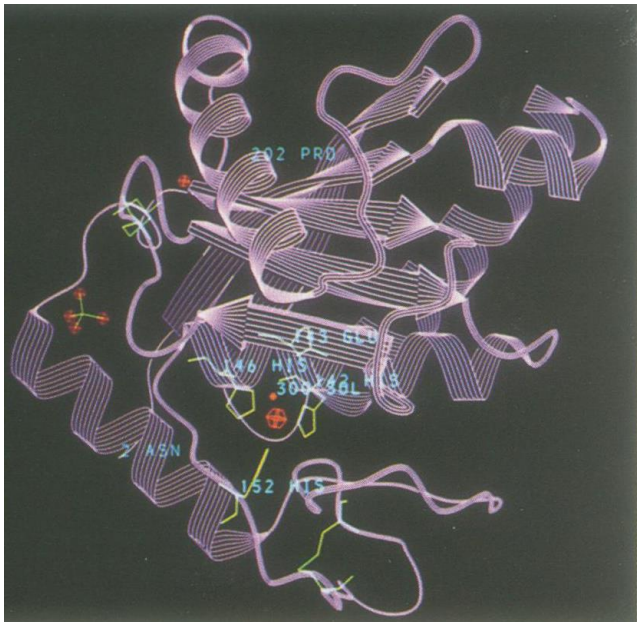


Figure 1.5. A ribbon model of Adamalysin II (P-I SVMP). An upper domain and a lower domain are separated by the active-site cleft which accommodates the zinc ion (red big sphere) co-ordinated by His 142, His 146 and His 152. A bound peptide substrate would presumably bind to the cleft running from left to right. The second red small sphere represents the bound calcium ion. Taken from Gomis-Ruth et al (1993).

1.2.2. The disintegrin domain

Disintegrins are small cysteine-rich proteins (49-90 amino acids) and typically possess an Arg-Gly-Asp (RGD) recognition sequence on an extended loop that can inhibit integrin-mediated platelet aggregation and cell-matrix interactions (84, 85). Kini and Evans proposed that metalloproteinases and disintegrins are derived by proteolysis from a common precursor and the sites of the proteolysis were determined by disulfide bridges, glycosylation and accessibility of interdomain cleavage sites (86). Snake venom disintegrins function as potent inhibitors of platelet aggregation by selectively blocking the binding of fibrinogen to integrin $\alpha_{IIb}\beta_3$ receptor, and the diverse integrin binding motifs act on different types of integrin receptors. The RGD motif blocks the $\alpha_8\beta_1$, $\alpha_5\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ integrins; MLD blocks the $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_3\beta_1$, $\alpha_6\beta_1$ and $\alpha_9\beta_1$ integrins; VGD and MGD block the $\alpha_5\beta_1$ integrin; KGD blocks the $\alpha_{IIb}\beta_3$ integrin; WGD blocks $\alpha_5\beta_1$, $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ integrins and KTS/RTS blocks $\alpha_1\beta_1$ integrin (85).

1.2.3. The disintegrin-like domain

The disintegrin-like domain in P-III SVMPS may function as a scaffold which spatially allocates the metalloproteinase domain and cysteine-rich domain rather than directly interacting with integrins. The disintegrin-like domain of P-III contains an XXCD (SECD) sequence instead of the typical RGD motif. The ECD region is structurally related to the RGD region of RGD disintegrins. Elucidation of the three-dimensional structure of bothropasin (P-III SVMPs) from *Bothrops jararaca* helped to reveal the mechanism of action of the multimodular proteins. The ECD motif is stabilized by the Cys227-Cys310 disulfide bond (between the disintegrin and cysteine-rich domains) and by one calcium ion. The side chain of Glu276 of the ECD motif is exposed to solvent and is free to establish chemical interactions. The differences in the disintegrin-like, cysteine-rich or disintegrin-like cysteine-rich domains may be influenced by target binding selection, which in turn can generate a substrate diversity or a specificity for the metalloproteinase domain (87). In VAP1 (vascular apoptosis-inducing protein-1), a P-III SVMP from *Crotalus atrox* venom, the DECD sequence containing disintegrin domain extends from the metalloproteinase domain and is opposed to the catalytic site. The disintegrin domain forms a C-shaped arm, together with the cysteine-rich domain, and its concave surfaces toward the metalloproteinase domain. Thus, the disintegrin-loop stabilizes the continuous structure and is inaccessible for protein binding (80).

1.2.4. The cysteine-rich domain

The cysteine-rich domain of P-III SVMPs has 2 subdomains: the wrist (C_w-domain, residues 488–505) and the hand (C_h-domain, residues 505–610). The C_h-domain, together with the C_w-domain, constitutes a novel fold (88). Jararhagin, a P-III SVMP isolated from *Bothrops jararaca* venom inhibits collagen-induced platelet aggregation. Jararhagin recognizes the platelet collagen receptor $\alpha_2\beta_1$ integrin and cleaves its β_1 subunit, resulting in loss of platelet responses to collagen (89). Interestingly, the binding of jararhagin to collagen and $\alpha_2\beta_1$ integrin occurs through two independent motifs, which are located on disintegrin-like and cysteine-rich domains, respectively (90).

1.2.5. The C-type lectin-like domain (C-type lectin-like proteins, CLPs)

The basic structures of CLPs include two homologous subunits: subunit α (A chain) of 14 to 15 kDa and subunit β (B chain) of 13 to 14 kDa. CLPs occur in a variety of oligomeric forms, including αβ, $(\alpha\beta)_2$, and $(\alpha\beta)_4$. The amino acid sequence of CLP subunits are 15–40% identical to the carbohydrate recognition domain (CRD) of C-type lectins, including mammalian mannose-binding protein (MBP) and snake venom galactose-specific lectin (91). It is now clear that the CLPs form a new protein family. CLPs possess various biological functions including anticoagulant-, procoagulant- and platelet-modulating activities (92). C-type lectin-like proteins are found in proteins associated with a proteinase domain in a multidomain structure: the factor X activator, RVV-X, from *V. russelli* venom and the prothrombin activator, carinactivase, from *E. carinatus* venom. Russell's viper (*Vipera russelli*) blood coagulation factor X activator (RVV-X), a heterodimeric metalloproteinase, contains a C-type lectin-like domain as its light chain. The two light chains of RVV-X recognize the Ca^{2+}-bound conformation of the Gla domain in factor X and subsequently activate factor X to factor Xa (29). Similarly, the regulatory subunit of carinactivase (2 C-type lectin-like domains) selectively recognizes the normal Ca^{2+}-bound conformation of prothrombin containing all of the Gla residues (93).

1.3. Snake venom metalloproteinases from Russell's vipers

Many snake venom metalloproteinases have been identified in the different Russell's viper species. SVMPs isolated from Russell's viper subspecies and studied are listed in Table 1.4. Details of Russell's viper SVMPs for which sequences have been obtained can be found in Table 1.5.

Among the SVMPs from Russell's viper, RVV-X is a potent blood coagulation factor X activator. It belongs to the P-IIId class SVMPs as it has two subunits held by an inter-subunit disulfide linkage: larger subunit (heavy chain) is a P-III metalloproteinase whereas the smaller subunit (light chain) is a C-type lectin-like protein with two chains (light chain A & light chain B) covalently linked together by an inter-chain disulfide bond. The three peptide chains are assembled into a hook-spanner-wrench configuration. The metalloproteinase and disintegrin-like domains of heavy chain

resemble a hook, and the rest of the molecule forms a handle. A disulfide bridge between the Cys389 of heavy chain and Cys133 of light chain A links the two chains. Light chain A and B are linked by a disulfide bond between Cys79 and Cys77 of the respective chains. The concaved cleft formed between the light chains is likely to be a primary capture site (spanner) for Factor X by binding to the gamma-carboxyglutamic acid-rich (Gla) domain of Factor X in the presence of Ca^{2+}. The hook portion of RVV-X not only exerts a catalytic effect but may also regulate the binding affinity between molecules as a wrench configuration together with hyper-variable-region (HVR) in C_h (94, 95).

Table 1.4. SVMPs isolated from Russell's viper subspecies

Name	Molecular weight (kDa)	Source	Accession no.	Biological function
RVV-X (*D. r. siamensis*)	79	Venom from Japan Snake Institute	-	Activation of factor X (29)
RVV-X (*D.r.russelli*)	93	Venom from Latoxan, Rosans, France	-	Activation of factor X (96)
RVV-X (*D. siamensis*)	93	Venom from Floras Island, Indonesia	Q7LZ61	Strongly binds to protein S, factor X and IX (97)
RVV-X (*D.r.siamensis*)	91	Venom from QSMI, Thailand	-	Renal failure through intravascular clotting in renal microvasculature (30)
VRR-73 (*D. r. russelli*)	73	Venom from Calcutta, Snake Park, India	-	Fibrinolytic proteinase with haemorrhagic and esterolytic properties (31)
VRH-1 (*V. r. russelli*)	22	Venom from Calcutta, Snake Park, India	-	Severe lung hemorrhage (32)
RVBCMP (*D. r. russelli*)	15	Venom from Calcutta, Snake Park, Kolkata	-	α-fibrinogenolytic procoagulant proteinase causes distinct liver haemorrhage (33)
Daborhagin-M Daborhagin-K	65 65	Venom from Myanmar and Kolkata (eastern India)	P0DJH5 B8K1W0	Severe dermal haemorrhage Daborhagin-M has hydrolysis effect on α-fibrinogen, fibronectin, type IV collagen (34)
MRV-75	75	Snake Farm, Myanmar Pharmaceutical Factory, Myanmar	-	Haemorrhagic, oedema inducing, myonecrotic, coagulant, fibrinolytic, fibrinogenolytic activities (26)

Table 1.5. List of SVMP sequences that have already identified in Russell's vipers

Name	Country of Origin	Length (Amino acids)	Accession no.
RVV-X-HC	Thailand (63)	539	-
RVV-X-LC1	Thailand (63)	146	-
RVV-X-HC	Indonesia (97)	619	Q7LZ61
RVV-X-LC1	Indonesia (97)	123	-
RVV-X-LC2	Indonesia (97)	134	-
RVV-X-HC	Venom from Japan Snake Institute (29)	429	AAB22477.1
RVV-X-LC1	Venom from Japan Snake Institute (29)	123	AAB22478.1
RVV-X-LC1	Eastern Russell's viper [Zhong, et al (2005)]	146	Q4PRD1.2|SLLC1_DABSI #
RVV-X-LC2	Eastern Russell's viper [Zhong, et al (2005)]	158	Q4PRD2.1|SLLC2_DABSI #
RVV-X-LC1	Myanmar [Tsai et al (2009)]	146	gb|ADK22820.1 #
RVV-X-LC2	Myanmar [Tsai et al (2009)]	158	gb|ADK22819.1 #
RVV-X-HC	Indian [Tsai et al (2009)]	619	ADJ67475.1 #
RVV-X-LC1	Indian [Tsai et al (2009)]	146	ADJ67474.1 #
RVV-X-LC2	Indian [Tsai et al (2009)]	158	ADJ67473.1 #
Chain A (RVV-X-HC)	Enzyme Research Laboratories Inc. (95)	427	gi|162329887| pdb|2E3X|A
Chain C (RVV-X-LC1)	Enzyme Research Laboratories Inc. (95)	122	gi|162329889| pdb|2E3X|C
Chain B (RVV-X-LC2)	Enzyme Research Laboratories Inc. (95)	134	gi|162329888| pdb|2E3X|B
RVV-X_HC	Sri Lanka (98)	10*	gi|298289237| sp|P86536.1|VM3CX_DABRR
RVV-X-LC2	Taiwan [Lin et al (2011)]	158	gi|380765752| gb|AFE61611.1|#
Russelysin (Daborhagin_K)	India (34)	615	gb|AAZ39880.1
Daborhagin-M	Myanmar (34)	25*	P0DJH5 (VM3DM_DABSI)

*Incomplete amino acid sequences

Direct submission to EMBL/GenBank/DDBJ databases

1.4. Natural inhibitors of snake venom metalloproteinases

Snakes show resistance to their own venom to protect themselves in a variety of ways. Natural snake venom metalloendopeptidases inhibitors (SVMPIs) have been isolated from snakes and mammals (99). The first natural SVMPI purified from sera of *Protobothrops* (Trimeresurus) *flavoviridis* is habu serum factor (HSF) (100). It is a 323-amino acid-long glycoprotein in possession of two cystatin-like domains. The HSF has a broad inhibitory specificity to several P-I, P-II and P-III classes SVMPs from venom of the same species (101). BJ46a from sera of *Bothrops jararaca* (102), cMSF from serum of the Chinese (*Gloydius blomhoffi brevicaudus*) and jMSF from serum of the Japanese (*G. blomhoffi)* mamushis (103) inhibited proteolytic activities of the P-III class SVMPs Jararhagin from *B. jararaca*; HR1A, HR1 B from *P. flavoviridis*; brevilysins H2, H3, H4, H6 from *G. blomhoffi brevicaudus* venom respectively. The BJ46a, cMSF and jMSF are homologous to HSF which is a member of the fetuin family, part of the cystatin superfamily. BJ46a showed a non-covalent complex with their metalloproteinase domains on gel filtration analysis. The exact mechanism of action involves an interaction of these inhibitors with metalloproteinases, but are still incompletely defined.

Secondly, DM40 and DM43 isolated from *Didelphis marsupialis* (common opossum) contain immunoglobulin-like domains and their combined fraction inhibited the proteolytic activity of *Bothrops jararaca* venom upon fibrinogen, fibrin, collagen IV, laminin, and fibronectin. These SVMP inhibitors from the immunoglobulin supergene family also form non-covalent complexes with jararhagin, the main hemorrhagic metalloproteinase from *B. jararaca* venom (104).

Thirdly, the antihaemorrhagic factor, erinancin isolated from *Erinaceus europeus* muscle is a member of the ficolin/opsonin P35 superfamily. It was shown to inhibit a metalloendopeptidase from *B. jararaca* venom. Erinacin showed a “flower bouquet”-like structure under electron microscopy, a similar structure found in the plasma molecules, ficolin. The proposed inhibitory mechanisms of the ericancin were: 1) the fibrinogen-like domain of erinacin could contribute to the inhibition of the metalloproteinase by recognizing the N-acetylglucosamine residue and 2) the collagen-like domain of erinacin could also exert an affinity for the haemorrhagic

metalloproteinase as a substrate mimic since SVMPs show activity towards collagen substrates (105) .

In addition to high molecular weight SVMPIs, snake venom of several species are known to be contain citrate and small peptides to protect against auto-digestion by SVMPs. They can bind selectively to SVMPs in the venom glands to protect gland of inhibition tissues and venom factors from self-digestion by SVMPs (106). Three endogenous peptides: pyroGlu-Lys-Trp (pEKW), pyroGlu-Asn-Trp (pENW) and pyroGlu-Gln-Trp (pEQW) isolated from venom of Taiwan habu (*Trimeresurus mucrosquamatus*) showed inhibitory action to proteolytic activity of metalloproteinases present in the crude venom (107). It has been reported that these peptide inhibitors regulate the proteolytic activities of their SVMPs in a reversible manner under physiological conditions. The binding of tripeptide inhibitors causes some of the residues around the inhibitor-binding environment of the metalloproteinase to slightly move away from the active-site center, and displaces two metal-coordinated water molecules by the C-terminal carboxylic group of the inhibitors. The three principal interactions that stabilize the binding of inhibitors to the metalloproteinase are: 1) The Trp indole ring of inhibitors is stacked against the imidazole ring of His143 in the S^{-1} pocket of the proteinase; 2) the middle residue of the tripeptide inhibitors are stabilized at the S^{-2} site of the proteinase by three possible hydrogen bonds; 3) the pyro-ring of these inhibitors is held at the S^{-3} site of the proteinase by hydrophobic interactions (108). Other pit vipers venom also contain endogenous tripeptides: pEQW and pENW in *Bothrops asper* (109) and rattlesnakes (110). African vipers such as *Echis ocellatus* and *Cerastes cerastes cerastes* have the peptides inhibitor pEKW. This tripeptide is encoded by tandemly-repeating elements from the transcripts which also contain a CNP (C-type natriuretic peptide) homologous sequence at the C-terminus (111). Two peptides: PtA (pENW) and PtB (pEQW) isolated from venom liquor of *Deinagkistrodon acutus* (Hundred-pacer viper) showed anti-human platelet aggregation activity *in vitro* and protection effects on ADP-induced paralysis and formation of pulmonary thrombosis in mice (110).

1.5. Aim and objectives

The clinical presentations from Russell's viper-bite patients vary among different countries, with severe renal failure, coagulopathy, oedema and shock mostly significant in Myanmar, among other countries. Even within the same snake subspecies (*D. r. siamensis*), the coagulopathy, oedema, haemorrhage and shock symptoms are more prominent in Myanmar envenoming, while haemolysis is markedly present in Thailand. This could reflect the high protease activity of Myanmar Russell's viper venom (26) whereas relatively high phospholipase A_2 activity that damaging red cell membrane present in Thailand Russell's viper venom (27). The specific protein families found in different snake families represent related proteins that vary in amino acid sequences and abundance and thereby contribute to the differences in the overall biological activity of individual venoms. The snake venom metalloproteinase, the major constituent in Viperidae venom, from Myanmar Russell's viper has not been characterised at the molecular level in terms of its composition and sequence.

Thus, the aim of this study was to study the SVMP toxins family in Myanmar Russell's viper. The objectives of this thesis are 1) to identify, quantify and characterise the SVMPs transcripts from cDNA library of Myanmar Russell's viper venom glands; 2) to isolate and characterise the SVMPs from Myanmar Russell's viper venom; and 3) to compare the SVMP toxin sequences profile of Myanmar Russell's viper with those of other vipers. The hypothesis is that the SVMPs toxin profile of Myanmar Russell's viper may show a distinct pattern to produce characteristic bleeding clinical manifestations. The last objective is 4) to identify and characterise the SVMP inhibitors from Myanmar RV venom.

Chapter 2. Materials and methods

2.1. Materials

The crude venom and venom glands from Myanmar's Russell's vipers (*Daboia russellii siamensis*) were obtained from the Snake Farm, No. 1, Myanmar Pharmaceutical Factory, Yangon, Myanmar. Venom was milked 3 days before dissection of the venom glands to stimulate the production of messenger RNAs (mRNAs) in the venom glands.

Venom gland transcriptomes of 3 animals (2 adult males and 1 female), were sequenced. Adult males weighed 740 g & 950 g with snout-to-vent lengths of 40 & 41 inches and total-lengths of 46 & 48 inches. The adult female had a weight of 1,490 g with a snout-to-vent length of 44 inches and a total-length of 49 inches. The habitat of the snakes is Kyun Gyan Gon Township, Yangon, Myanmar. The experimental plan was approved by the Animal Care and Use Committee, Chulalongkorn University (CU-ACUC) (No. 17/2558). The venom glands were dissected while the snake was anaesthetised with chloroform. The tissues were kept in RNAlater solution (Ambion Inc., Canada, USA) to stabilise and protect the cellular RNA for storage without endangering the quality and quantity of RNA. The tissues samples were stored at -80°C until further processing.

2.2. Methods

2.2.1. Complementary DNA (cDNA) library construction

2.2.1.1. Preparation of samples for cDNA library construction

Total RNA from male gland tissues was extracted by using Trizol reagent (Life Technologies, Carlsbad, California) (112) and that from females was extracted by using a Total RNA Purification Kit (Jena Bioscience GmbH, Jena, Germany). The concentration of total RNA was determined with NanoDrop spectrophotomer (Thermo Fisher Scientific, Wilmington, USA). Later, the sample was stored at -80°C. Subsequently, mRNA from 2 male samples was isolated using PolyAT Tract mRNA Isolation System (Promega, Madison, USA). For the female sample, mRNA was isolated using a FastTrack MAG mRNA Isolation Kit (Invitrogen, Carlsbad, California). Both systems use MagneSphere® technology. The concentration of mRNA was then determined using a NanoDrop spectrophotometer. Then, mRNA was precipitated, concentrated using 3 M sodium acetate and isopropanol and stored at -80°C.

2.2.1.2. cDNA library construction

The CloneMiner™ II cDNA Library Construction Kit was used to construct the cDNA library of the snake venom gland. The basis of the methods is as follows. The single-stranded mRNA was converted into double-stranded cDNA using SuperScript® III Reverse Transcriptase. The attB1 adapter was ligated to the 5'end of double-stranded cDNA containing attB2-Biotin at the 3'end. After ligation, cDNAs were size-fractionated through Sephacryl® S-500 HR resin mini-column to remove residual adapters, primers and other small truncated cDNAs (<500 bp). The *att*B-flanked cDNA was then recombined into the pDONR™ 222 plasmid containing *ccd*B gene which is responsible for negative selection of vector in *E.coli* following recombination and transformation. Gateway® BP Clonase® II enzyme mix facilitate the recombination of *att*B-flanked cDNA with an *att*P-containing vector to create an *att*L-containing entry clone in BP recombination reaction. The entry clone was transformed into *E. coli*, ElectroMAX™ DH10B™ T1 Phage Resistant Cells, using the BioRad Gene Pulser.

Day 1, Synthesis of the first strand: 8 µg of mRNA was used for the priming reaction. The 2.0 kb RNA (2 µg), supplied with the kit, was used as a control. One microliter of biotin-*att*B2-oligo(dT) primer (30 pmol/L) was added and the mixture was incubated at 70°C for 7 min and gradually cooled to 45°C for over 15 min in the thermocycler. After that, 5 x First strand buffer (4µl), 0.1 M DTT (2µl), 10 mM dNTPs (1µl) and SuperScript III Reverse Transcriptase (4µl of 200 U/µl) were added into the priming reaction tube. The reaction tube was incubated in stepwise increments as follows:

45°C for 20 min,

50°C for 20 min then

55°C for 20 min

Synthesis of the second strand: The second strand was synthesized by pipetting the 5 x Second strand buffer (30 µl), 10 mM dNTPs (3 µl), *E.coli* DNA ligase (10 U/ µl, 1 µl), *E.coli* DNA polymerase I (10 U/ µl, 4 µl), *E.coli* RNase H (2 U/ µl, 1 µl) into the first strand reaction tube and the mixture was incubated at 16°C for 2 h. To create blunt-ended cDNA, 2 µl of T4 DNA polymerase was added and then incubated at 16°C for 5 min.

Phenol/Chloroform extraction: The cDNA was extracted by adding 160 µl of phenol: choloroform: isoamyl alcohol (25:24:1). After centrifugation the sample at 16,000 x g at 25°C for 5 min, the aqueous upper phase was transferred into a new 1.5 ml tube.

Ethanol precipitation: Into the above tubes (both sample and control RNA tubes), glycerol (20 µg/ µl, 1 µl), 7.5 M NH_4OAc (80 µl) and 100% ethanol (600 µl) were added and the tubes were placed at -80°C for 10 min. The supernatant was removed after centrifugation at 16,000 x g at 4°C for 25 min. Next, the cDNA was washed twice with 150 µl of 70% ethanol. The cDNA pellet obtained was air-dried at room temperature (RT) for 10 min. The pellet was resuspended in 22 µl of DEPC-H_2O before adapter ligation.

Adapter Ligation: The cDNA pellet was placed on ice and into which 5 x Adapter Buffer (10 µl), attB1 Adapter (1 µg/ µl, 4 µl), 0.1 M DTT (8 µl) and T4 DNA ligase (1 U/ µl, 6 µl) were added. The mixture was gently mixed and incubated at 16°C for 16 h.

Day 2, Size Fractionation of cDNA: Before fractionation, the ligation reaction was stopped by incubation at 70°C for 10 min. The tubes were placed on ices while the purification column was set up. Firstly, the Sephacryl® S-500 HR resin column was washed with 0.8 ml of TEN buffer (0.1 M Tris-HCl, pH 8.0; 0.01 M EDTA, pH 0.8 & 1 M NaCl) 4 times. The entire sample was purified through the mini-column at a flow rate of 25µl per min. The fractions were collected into three 1.5 ml microfuge tubes (E1, E2 and, E3). Tubes E2 and E3 contained cDNA and E2 contained larger-sized cDNA than E3 tube. Next, the column was washed with 3 ml of TEN buffer and the effluent was collected into W1, W2, and W3 tubes. The quantity of cDNA was measured using NanoDrop spectrophotometry. The collection tubes were kept at -80°C.

Day 3, BP recombination reaction: BP recombination is the site-specific recombination of *att*B-flanked cDNA to an *att*P-containing donor vector. The purified cDNA was precipitated, washed and resuspended. The final yield of cDNA was determined using NanoDrop spectrophotometry. The ratio of cDNA insert to vector was 1:4 and the reaction mixtures were prepared as follows (Table 2.1).

Table 2.1. Reaction volume for BP recombination reaction

	Sample	RNA Control	BP Controls	
			Negative Control	Positive Control
*att*B-flanked cDNA	3 µl	3 µl	-	-
pDONR^{TM} 222(150 ng/ µl)	2 µl	1.6 µl	1.6 µl	1.6 µl
pEXP7-tet (50ng/ µl)	-	-	-	0.5 µl
TE buffer, pH 8.0	2 µl	2.4 µl	5.4 µl	4.9 µl
BP Clonase II enzyme mix	3 µl	3 µl	3 µl	3 µl
	10 µl	10 µl	10 µl	10 µl

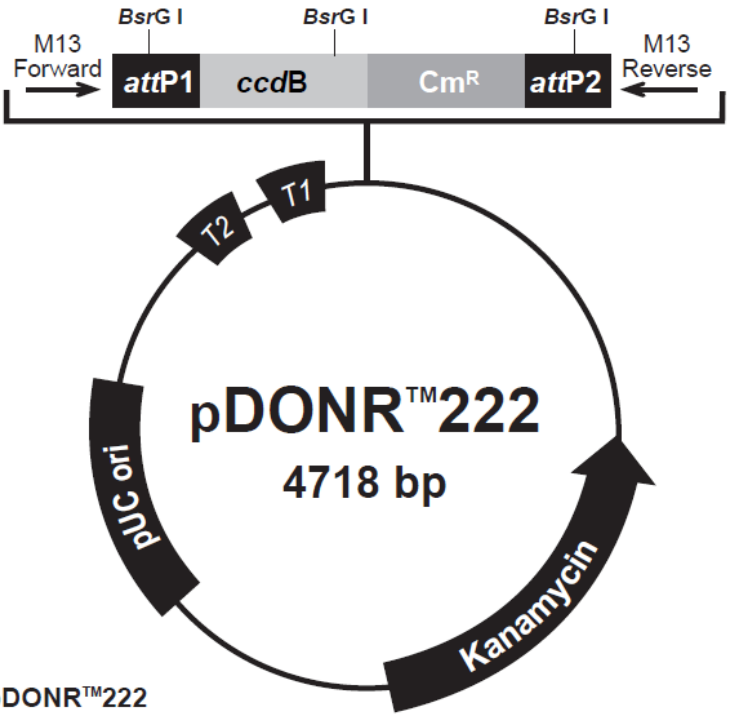
The mixture and all the reagents were placed on ice and the final mix was incubated at 25°C for 20 h. The vector pDONR^{TM} 222 is a 4718 bp, kanamycin-resistance Gateway entry vector. This entry vector can shuttle the cDNA insert into a Gateway destination vector for further gene expression and functional analysis. The map of the pDONR^{TM} 222 is described in Figure 2.1.

pEXP7-tet control DNA was used as a positive control. pEXP7-tet is an approximately 1.4 kb fragment consisting of the tetracycline resistance gene and its promoter (Tc^{r})flanked by *att*B sites. The pEXP7-tet in pDONR^{TM}222 vector results in the entry clones expressing both tetracycline- and kanamycin-resistant genes.

Day 4, Preparation for transformation: The recombination reaction was stopped by adding 2 µl of Proteinase K and the reaction tubes were incubated at 37°C for 15 min, then at 75°C for 10 min. After that, sterile water was added to the BP recombination mix to a final volume of 100 µl. The cDNA was precipitated using 100% ethanol at -20°C overnight.

Preparation of agar plates: A total of 26 agar plates was used: 6 plates were prepared with 25 ml of LB agar solution in a 10 cm petri-dish containing 50 µg/ml kanamycin for each cDNA sample, RNA control, BP negative control and BP positive control, as well as 2 plates containing 100µg/ml ampicillin for pUC19 transformation

control. After solidification, the agar plates were stored in 4°C in the inverted position.



Comments for pDONR™222
4718 nucleotides

*rrn*B T2 transcription termination sequence: bases 58-85 (c)
rrnB T1 transcription termination sequence: bases 217-260 (c)
M13 Forward (-20) priming site: bases 327-342
*att*P1: bases 360-591
*Bsr*G I restriction sites: bases 442, 1232, 2689
*ccd*B gene: bases 987-1292 (c)
Chloramphenicol resistance gene: bases 1612-2295 (c)
*att*P2: bases 2543-2774 (c)
M13 Reverse priming site: bases 2816-2832
Kanamycin resistance gene: bases 2899-3714 (c)
pUC origin: bases 4045-4718
(c) = complementary strand

Figure 2.1. Elements of pDONR™ 222 vector. Retrieved from "http://tools.lifetechnologies.com/content/sfs/vectors/pdonr222_map.pdf . "

Day 5, Transformation**:** The cDNA sample was washed in 150 µl of 70% ethanol and resuspended in 9 µl of TE buffer. The sample was divided into 6 aliquots each

containing 1.5 µl of cDNA sample. The ElectroMAX cells were thawed on ice. The amounts of cells used in transformation are described in Table 2.2.

Table 2.2. Reaction volume for transformation reaction

	pUC 19 Control (10 pg/µl)	BP Negative Control	BP Positive Control	RNA Control	Sample
Amount	1 µl	1.5 µl	1.5 µl	1.5 µl	1.5 µl
ElectroMAX cells	50 µl	50 µl	50 µl	50 µl	50 µl
Number of aliquots	1	2	2	2	6

Each mixture was mixed gently by pipetting without introducing air bubbles. The mixtures were then transferred into a cold 0.1 cm cuvette. The electroporation condition was 2.2 kV, 200 Ω, 25 µF. The electroporated cells were diluted with 1 ml of SOC medium and transferred into a 15 ml Falcon tube. The transformed cells solution was shaken for 1 hour at 37°C at 250 rpm to allow expression of the kanamycin resistance marker. All electroporated cells representing one library were pooled into a 15 ml tube. The cDNA libraries were stored in an equal volume of sterile freezing media (6 ml for sample, 2 ml for each RNA control, BP negative control, BP positive control, but not to pUC 19 control) at -80°C.

Plating Assay: The sample aliquots were diluted with SOC medium as shown below.

	pUC 19 Control	BP Negative Control	BP Positive Control	RNA Control	Sample
Dilutions	10^{-2} - -	Undiluted 10^{-1} 10^{-2}	10^{-2} 10^{-3} 10^{-4}	10^{-2} 10^{-3} 10^{-4}	10^{-2} 10^{-3} 10^{-4}

For each 1:10 serial dilution, 100 µl of the sample was diluted with 900 µl of SOC medium. Then, 100 µl of each dilution was spread onto a pre-warmed LB agar plate

containing the appropriate antibiotic. The duplicated assay plates were incubated overnight at 37°C.

Day 6, Characterisation of cDNA libraries: The colonies from each LB plate were counted. The titre for each plate was calculated by the following equation.

$$cfu/mL = \frac{colonies\ on\ plates \times dilution\ factor}{volume\ plated\ (mL)}$$

The titre for each plate was used to calculate the average titre of the cDNA library. The total colony-forming units were determined by multiplying the average titre by the total volume of the cDNA library.

Day 7, Qualification of the cDNA library: The average insert size and percentage of recombinants were measured to characterise the cDNA library. Each colony of the clones was picked up with a sterile metal loop from the assay plate and cultured overnight in 3 ml of LB medium containing 50µg/ml kanamycin. The isolation of plasmid DNA was performed by using a High Pure Plasmid Isolation Kit (Roche Applied Science). The purified plasmid DNA (500 ng) was digested with 5 units of *Bsr*G I at 37°C for 1 h. The insert cDNA size was estimated using a 1% agarose gel stained with ethidium bromide.

2.2.1.3. DNA sequencing

For each DNA sequencing reaction, 200 ng of template DNA, corresponding to a single colony carrying an insert cDNA, was used. The PCR reaction was carried out in a 20µl reaction containing 4 µl of terminator-ready reaction mix from a BigDye™ Terminator v3.1 Ready Reaction Cycle Sequencing kit, 40 pmol of the universal M13 forward primer (5'-TGTAAAACGACGGCCAGT-3') and DNA template. After incubation at 96°C for 1 min, amplification was carried out for 25 cycles of 96°C for 10 s of denaturation; 50°C for 10 s of annealing and 60°C for 75 s of extension. The duration of extension (T_e) was increased in the last 10 cycles. T_e was extended by 15 s for cycle 16-20 and further extended an additional 30 s for cycles 21-25. This stepped T_e protocol increases the signal strength and thus less fluorescent terminator mix was required

(113). The DNA was then precipitated by the addition of 2 µl of 3 M sodium acetate, pH 5.8 and 50 µl of 95% ethanol and incubated at -20°C overnight. After centrifugation at 8,000 x g for 20 min, the pellet was washed with 1 ml of 70% ethanol and dried in a heat-block at 95°C for 2 min. The PCR-amplified products were then sequenced in a 3130 Genetic Analyzer, Life Technologies, USA.

2.2.1.4. Annotation of sequences

The DNA sequence chromatograms were analysed and edited with Finch TV (version 1.4.0, Jim Patterson, Benjamin Chamberlain & Diana Thayer, Geospiza Inc., Seattle, WA, USA) software. The sequences with low quality values at both ends of the sequence were firstly trimmed. After removing the vector sequence with VecScreen and adaptor sequences, the edited sequence was searched against the NCBI GenBank database using Blastn algorithm to identify similar sequences with an e-value cutoff $<10^{-5}$.

2.2.2. Next-Generation sequencing, *de novo* assembly and, sequence analysis

2.2.2.1. Next-Generation sequencing (NGS)

The mRNA (details in 2.2.1.1.) was resuspended in RNase-free water before sending to Macrogen Inc. Geumchun-gu, Seoul, Korea for RNA sequencing. The mRNA sequencing libraries were prepared using the TruSeq RNA sample preparation kit (Illumina) with selected insert sizes of 200-400 bp. Libraries were then sequenced on the Illumina HiSeq2000 platform using (2x100bp) paired-end reads.

2.2.2.2. *De novo* assembly

NGS reads were quality assessed with the quality assessment software FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The raw data from Illumina platform was then processed by Trimmomatric (0.32) (http://www.usadellab.org/cms/?page=trimmomatic) to remove adaptor sequences, error nucleotides and low quality sequences. The resulting sequencing reads were assembled using Trinity (r20140717) (http://trinityrnaseq.sourceforge.net/) and CLC Genomics WorkBench 7.5 (*de novo* assembly) (CLC bio, a QIAGEN Company, Aarhus, Denmark). Then, QUAST tool was used to assess the performance of 2 assembly software.

2.2.2.3. Annotation and prediction of ORF/CDS

De novo assembled transcripts were annotated in three ways. Firstly, Blastx (Gene Ontology) was used against a protein sequence database (go_v20140820). Secondly, transcript annotation was achieved using Blastn search against the filtered NCBI nucleotide sequences database with two-keyword search "venom" and "serpents". The annotation with high scores (with an e-value cutoff of $<10^{-5}$) at the top of the Blastn hit list were considered after filtering out genome sequences, transcribed sequences and duplicated contigs out.

Matches to known toxins were then grouped manually in a Microsoft Excel work sheet according to key words of known toxin names. From different toxin groups, metalloproteinase contigs were fetched from whole data set using Python programming language. Lastly, using Blastx with the cut-off e-value of 1.0 e^{-5}, analysis of homologous was carried out against the NCBI SwissProt database. The best annotation was taken after Clustal Omega alignment of the translated sequences.

The CDS (Coding sequences) were selected based on Blastn and Blastx results. The open reading frame of each sequence was predicted by ORF Finder or OrfPredictor and their conserved domains were identified against NCBI's Conserved Domain database.

2.2.2.4. Transcript abundance analyses

Transcript quantification of the *de novo* assembly was carried with RSEM (1.2.15), which estimated transcript abundance based on the metrics standardized the number of RNA-seq reads mapped to a particular exon by the total number of mapped reads and the size of the exon. RSEM calculates maximum likelihood abundance estimate as well as posterior mean estimates and 95% credibility intervals for genes/isoforms. It enables accurate transcript quantification for species without sequenced genome (114) FPKM (Fragments Per Kilobase of exon per Million fragments Mapped) values were used for analysis of relative transcript expression level.

2.2.3. Cloning of snake venom metalloproteinases (SVMPs)

Depending on the sequence information achieved from NGS data, amplification of mRNA transcripts was performed by either 5'RACE for transcripts with incomplete coding regions or conventional PCR for transcripts with complete coding regions.

2.2.3.1. 5'RACE

The 5'RACE (Rapid Amplification of cDNA Ends) system (Invitrogen, Carlsbad, California) is a set of reagents for 1) synthesis of first strand cDNA, 2) purification of first strand products, 3) homopolymeric tailing, 4) amplification of taget cDNA by PCR. The procedure is summarized in Figure 2.2.

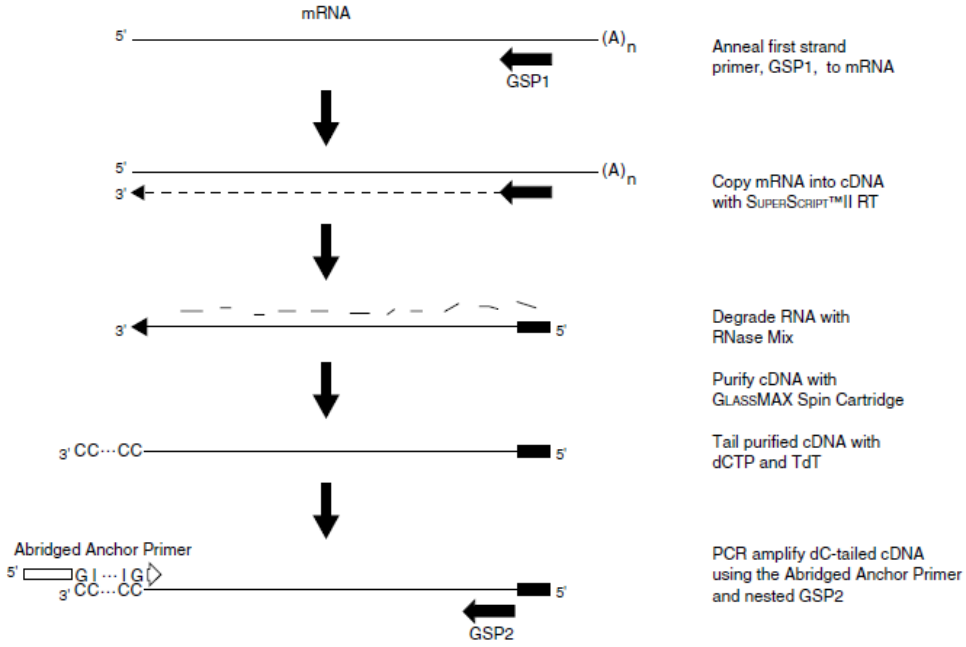


Figure 2.2. Overview of the 5'RACE procedure.

GSP1 = Gene specific primer 1. GSP2 = Gene specific primer 2**.** Retrieved from "http://download.bioon.com.cn/upload/month_0901/20090112_c9df111de76ffd695b35LOZnT25dbdZr.attach.pdf"

1) Synthesis of first strand cDNA

For amplification of nucleic acid sequences from a messenger template between a defined internal site and unknown sequences at the 5'-end of the mRNA, a minimum of two antisense gene-specific primers are required (115). GSP1 should be designed to anneal at least 300 bp from the mRNA 5'-end. The location of GSP2 should be at

further upstream of GSP1 within the cDNA product. First strand cDNA synthesis is primed using GSP1 with SuperScript™ II reverse transcriptase.

In a 0.5 ml microcentrifuge tube, 500 ng of sample mRNA was mixed with 2.5 pmole of GSP1 in a reaction volume of 15.5 µL. The mixture was incubated at 70°C for 10 min to denature RNA and then chilled on ice for 1 min. The following were added after the tube was briefly centrifuged: 10X PCR buffer (2.5 µL), 25 mM $MgCl_2$ (2.5 µL), 10 mM dNTP mix (1µL) and 0.1 M DTT (2.5µL). The mixture was then mixed gently, briefly centrifuged and incubated at 42°C for 1 min. Next, 1 µL of SuperScript™ II RT was added and incubated at 42°C for 50 min to synthesize the cDNA.

After 1st strand cDNA synthesis, the original mRNA template was removed by treatment with RNase Mix (mixture of RNase H and RNase T1). RNase H degrades cDNA: RNA hybrid and RNase T1 degrades single-stranded RNAs. 1 µL of RNase mix was added into the synthesized cDNA tube and mixed thoroughly. The reaction mixture was then incubated at 37°C for 30 min.

2) Purification of first strand products

The unincorporated dNTPs, GSP1 and proteins were separated from cDNA using a S.N.A.P column. Firstly, 120 µL 6 M NaI binding solution (equilibrated at room temperature) was added to the first strand reaction. Then, the cDNA/NaI solution was transferred to a S.N.A.P column and centrifuged at 13,000 x g for 20 s. The first flowthrough was kept until recovery of the cDNA was confirmed by nanodrop. Next, 0.4 mL of cold (4°C) 1x wash buffer was added to the spin cartridge and centrifuged at 13,000 x g for 20 s. The flowthrough was discarded. This wash step was repeated three times. Later, the cartridge was washed with 400 µL of cold (4°C) 70% ethanol as above. After removing the final 70% ethanol wash from the tube, the cartridge was centrifuged at 13,000 x g for 1 min. Finally, the cartridge was transferred into a sample recovery tube and 50 µL of preheated (65°C) sterilised, distilled water was added into the cartridge tube. The cDNA was eluted by centrifugation at 13,000 x g for 20s.

3) Homopolymeric tailing

A homopolymeric tail (CC…CC) was then added to 3'end of the cDNA using TdT (Terminal deoxynucleotidyl Transferase) and dCTP. Firstly, the following components were added into 3 separate tubes (Table 2.3).

Table 2.3. Reaction volume for homopolymeric tailing

Components	cDNA (50 ng) with dC-tail	cDNA (100 ng) with dC-tail	cDNA without dC-tail
DEPC-treated water	14.9	13.4	14.9
5X tailing buffer	5.0	5.0	5.0
2 mM dCTP	2.5	2.5	2.5
S.N.A.P.-purified cDNA sample (32 ng/µL)	1.6	3.1	1.6
Final volume (µL)	24.0	24.0	24.0

The above mixtures were incubated at 94°C for 3 min and then chilled on ice for 1 min. Then, the tubes were centrifuged briefly and placed on ice. Subsequently, 1 µL of TdT was added to the first two tubes and mixed gently the mixtures. All 3 tubes were incubated at 37°C for 10 min.

After this, the TdT was inactivated at 65°C for 10 min. The contents of the reaction were collected by brief centrifugation and the tubes were placed on ice.

4) PCR amplification of target cDNA

The dC-tailed cDNA was then amplified by PCR using a nested GSP2 which anneals to 3' to GSP1 and Abridged Anchor Primer (AAP) provided with the system. The AAP contains poly G sequence which is complementary to dC-tailed region of the cDNA and acts as the sense primer in the PCR reaction. In three 0.5 ml thin-wall PCR tubes sitting on ice, the following reagents were added (Table 2.4).

Table 2.4. Reaction volume for PCR amplification of target cDNA

Components	cDNA (50 ng) with dC-tail	cDNA (100 ng) with dC-tail	cDNA without dC-tail
Sterilized, distilled water	33.5	31.6	30.6
10X PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl]	5.0	5.0	5.0
25 mM $MgCl_2$	3.0	3.0	3.0
10 mM dNTP mix	1.0	1.0	1.0
Nested GSP2 (10 µM)	2.0	2.0	2.0
Abridged Anchor Primer (AAP)	2.0	2.0	2.0
Sample (300 ng)	3.0	4.9	5.9
Taq DNA polymerase (5 units/µL)	0.5	0.5	0.5
Final volume	50.0	50.0	50.0

The mixture was mixed and the PCR condition was set depending on the individual gene.

2.2.3.2. PCR amplification of cDNA

First strand cDNA was synthesized from total RNA with Oligo(dT)18 primer and reverse transcriptase from Revert Aid H Minus First Strand cDNA Synthesis kit (Thermo Scientific, Carlsbad, California). The cDNA was then used as a template in PCR amplification of individual genes with gene specific forward and reverse primers using FastStart Taq DNA Polymerase, dNTPack (Roche Diagnostics GmbH, Mannheim, Germany). A sense primer complementing the N-terminal residues just before start codon and an antisense primer beyond the stop codon were designed according to the ORF of full-length SVMP transcripts obtained from NGS. PCR conditions were set according to primers designed for the individual transcript, such as annealing temperature 5°C below Tm of the primers, the extension time of approximately 1 min per kilobases of expected products. The sizes of PCR products were identified on agarose gel electrophoresis.

2.2.3.3. Cloning of PCR products

The PCR product was purified using a Wizard® SV Gel and PCR Clean Up System (Promega, Madison, USA). Next, the purified PCR product was ligated into pGEM-T Easy vector (Promega, Madison, USA). The vector was then transformed into JM109 competent cells (Promega, Madison, USA) or into Stellar™ competent cells (Clontech, Mountain View, California). The white colonies were later selected from agar plates containing IPTG/X-Gal/Amplicillin and cultured in the LB medium containing ampicillin in shaking incubator 37˚C at 225 rpm overnight. The plasmids were then isolated with GF-1 Plasmid DNA Isolation kit (Vivantis, Subang Jaya, Malaysia). The size of cDNA insert was later checked by agarose gel electrophoresis after digestion with EcoRI (Promega, Madison, USA) and/or Not I (Vivantis, Subang Jaya, Malaysia) restriction enzymes. Finally, the plasmids with correct cDNA inserts were used for DNA sequencing.

2.2.3.4. DNA sequencing

DNA sequencing was done by using BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster, Califonia) on a TGradient thermocycler (Biometra, Gottingen, Germany). Then the precipitated DNA sample were analyzed with 3120 Genetic Analyzer (Applied Biosystems, CA, USA). For some samples, the plasmids were prepared according to protocol from Macrogen and sent to Macrogen (Macrogen Inc. Geumchun-gu, Seoul, Korea) for sequencing. The plasmids containing individual genes were firstly used for DNA sequencing with pUC/ M13, Forward (24 mers) and pUC/ M13, Reverse (22 mers) primers (Promega, Madison, USA). The walking primers were then designed depending on the contig sequences if the candidate genes were longer than 800 bp. Walking primers are the known complementary sequences located in equal interval (approximately 500 bp) of long gene sequence (1.3 to 7 kilobases).

2.2.3.5. Sequence analysis

The sequencing chromatograms were edited using FinchTV Version 1.4.0 (Geospiza Inc., Seattle, WA, USA). The corrected sequences were then aligned with original contig sequences from NGS. The DNA sequencing with walking primers was done for each clone. The DNA sequence reads were aligned and the whole consensus sequence was identified by using Geneious® 7.1.9 (Restricted, Biomatters Ltd, Auckland, New Zealand). The complete DNA sequence was aligned with original contig sequence again to check the similarity using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The translated amino acid sequences were deduced by using ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) and also aligned with that of original contig using Clustal Omega.

2.2.4. Purification of SVMPs and tripeptides

2.2.4.1. Protein concentration

Protein concentration was determined by using Bradford reagent (BioRad). The absorbance was measured at 595 nm and the calibration curve was prepared with a bovine gamma globulin standard (0-1.5 mg/mL).

2.2.4.2. Purification of SVMPs

All chromatographic procedures were performed on either a Bio-Cad Vision Workstation (GFC) or a GE Healthcare AKTA System (anion-exchange, HIC and RP-HPLC)

2.2.4.2.1. Gel filtration chromatography (GFC)

Crude venom [100 mg] dissolved in 5 mL of 0.01 M phosphate buffered saline (pH 7.4) was applied to a Superdex 200 column (5 x 160 cm) pre-equilibrated with the same buffer. Elution was carried out with the same buffer. The flow rate was 2 mL/min and 6 mL fractions were collected. The fractions having metalloprotease activity (section 2.2.5.) (fractions 15-18) were combined for further purification. The fractions eluting near the total volume were analysed for tripeptides using RP-HPLC with subsequent MS analysis.

2.2.4.2.2. Anion-exchange chromatography

The SVMP-containing sample obtained from GFC was applied to a Resource Q anion-exchange column (6 mL) pre-equilibrated with 0.05 M Tris-Cl buffer (pH 8.0). Elution was achieved with a linear NaCl gradient from 0 to 0.5 M in the same buffer at a flow rate of 0.6 mL/min and 1.8 mL fractions were collected. Elution was monitored at 280 nm.

2.2.4.2.3. Hydrophobic interaction chromatography (HIC)

To further purify the SVMPs for protease activity measurements, fractions from Resource Q were loaded onto a Phenyl Superose column (1 mL) equilibrated in 2.5 M NaCl, 50 mM Tris-Cl, pH 7.8. Samples in Tris-Cl were adjusted to 2.5 M in NaCl and were centrifuged at 10,000 x g for 5 min before loading onto the column. Separation was achieved by a 30 minute-gradient of 2.5 - 0 M NaCl in 50 mM Tris-Cl, pH 7.0, using a flow rate of 0.25 mL/min. Elution was monitored at 280 nm and 0.25 mL fractions were collected.

2.2.4.2.4. Reversed phase high performance liquid chromatography (RP-HPLC)

For mass spectrometry (MS) analysis, SVMPs were purified using RP-HPLC rather than HIC. Fractions from Resource Q chromatography were made up to 0.2% (v/v) in trifluoroacetic acid (TFA), centrifuged at 10,000 x g for 5 min and then applied to Phenomenex Aeris C4 column (150 x 2.1 mm, 5 micron). The proteins were separated in a two-part acetonitrile gradient in 0.08% TFA: 0-40% over 25 min then 40-65% over 5 min and elution was monitored at 280 nm. The flow rate was 0.15 mL/min and 0.25 mL fractions were collected.

2.2.4.3. Purification of SVMP inhibitors

Fractions from GFC suspected to contain small molecular weight components were made up to 0.2% (v/v) in TFA, centrifuged at 10,000 x g for 5 minutes and applied to Phenomenex Luna C18 RP-HPLC column (100 x 2.1 mm) equilibrated in 0.08% TFA. The components were separated at 0.15 mL/min with a three-part acetonitrile

gradient in 0.08% TFA: 0-12% over 5 minutes, 12-28% over 50 minutes and then 28-65% over 10 min. Elution was monitored at 280 nm.

2.2.4.4. Mass spectrometric analyses

Putative RVV-X (10 μg of R1 from RP-HPLC) was reduced, treated with iodoacetamide and digested with 1.0 μg trypsin in the presence of 2 M urea. The resulting peptides were analysed by LC-ESI-MS/MS using an Acquity UPLC CSH Peptide C18 RP column (Waters) connected to a Q-Exactive (Thermofisher) MS instrument. Peaks Studio 8.0 (BSI, Canada) was used to analyse the resulting MS/MS data against the sequences for Eastern Russell's viper RVV-X H chain VM3CX_DABSI (Q7LZ61) and light chains LC1 SLLC1_DABSI (Q4PRD1) and LC2 SLLC2_DABSI (Q4PRD2).

Putative daborhagin (5 μg of R2 from RP-HPLC) was reduced, treated with iodoacetamide and digested with 0.5 μg trypsin in the presence of 2 M urea. The resulting peptides were desalted and mass spectrometric analysis was performed using a MALDI-TOF instrument (Waters-Micromass). Samples were analysed by mixing a 1 μL solution of the tryptic peptides with an equal volume of 5.7 mg/mL α-cyano-4-hydroxycinnamic acid in 60% acetonitrile/0.1 % TFA and laying this onto a dried bed of 1 μL of 25 mg/mL α-cyano-4-hydroxycinnamic acid. Laser energy was set at 25% and detector voltage 1800 V. Ion spectra were collected in the mass range of 1,000-3,000 Da. Data analysis was performed using MassLynx. The tryptic peptide masses obtained were matched manually with those predicted (using ExPASy Peptide Mass) of a sequence for Daborhagin-K (VM3DK_DABRR; B8K1W0) retrieved from UniprotKB (www.uniprot.org) using search word 'Daborhagin'.

The purified tripeptides were also analysed with ESI-MSand ESI-MS/MS in the same way as Putative RVV-X without prior trypsin treatment.

2.2.4.5. Analysis by SDS-PAGE

Protein purity was determined by SDS-PAGE (116) on a 12 or 15% resolving gel and 4% stacking gel using a Mini-PROTEAN 3 electrophoresis system (BioRad). Samples were loaded in either reduced or non-reduced form. Gels were run at 200 V, 30 mA per gel, for 50 min. Proteins were visualised with Coomassie Brilliant Blue R250 V followed by destaining with methanol: water: acetic acid (30:60:10). Alternatively,

proteins were visualised by silver staining as performed by the method of Heukeshoven & Dernick (117).

2.2.5. Characterisation of SVMPs

2.2.5.1. Caseinolytic activity

The proteolytic activity was estimated by hydrolysis of heated casein using the Anson method (118). The reaction mixture, consisting of 500 µL casein (20 mg/mL) in 0.1 M Tris-Cl (pH 8.0), 20 µL venom was incubated for 30 min at 37˚C. The reaction was quenched by the addition of 500 µL of 5% trichloroacetic acid (TCA) at room temperature. After centrifugation at 10,000 x g for 5 min, the protein content of the hydrolysed substrate that was not precipitated with TCA, was determined by the Folin-Ciocalteau method (119). Thus, 400 µL of the supernatant was mixed with 1 mL of 0.5 M Na_2CO_3 and 200 µL of diluted (1:5) Folin-Ciocalteau's phenol reagent. The mixture was then incubated at 37°C for 30 min and the absorbance was measured at 660 nm. One enzyme unit is defined as the amount of enzyme which hydrolyses casein to produce color equivalent to 1.0 µmole of tyrosine per minute at pH 8.0 at 37°C.

2.2.5.2. Gelatinolytic activity

The gelatinolytic activity of the purified enzyme was analysed by zymography (120). The purified metalloproteinase was diluted in SDS sample buffer under non-reducing conditions and run on 10% SDS-polyacrylamide gels (0.75 mm) co-polymerized with 0.5 mg/mL of gelatin. After electrophoresis, the gels were washed in 2.5% Triton X-100 for 30 min and then washed three times in distilled water to remove any Triton. Gels were then incubated in developing buffer (50 mM Tris-Cl, pH 7.8, 200 mM NaCl, 5 mM $CaCl_2$, 0.02% Brij 35) for 18 h at 37˚C. The gels were stained with 0.5% Coomassie blue R-250 in methanol: acetic acid: water (5:10:85) solution and subsequently destained in methanol: acetic acid: water (10:5:85). The presence of gelatinolytic activity was defined as clear bands on the dark blue background.

2.2.5.3. Fibrinogenolytic activity

The fibrinogenolytic activity was assayed by SDS-PAGE (4% stacking/12% resolving gel) as described by Ouyang & Teng (121). Equal volumes of fibrinogen (1 mg/mL in 0.05 M Tris-Cl, pH 8.5) and 20 µg/mL of enzyme were incubated at 37°C for various times intervals. At 0, 5, 15, 30, 60 and 120 min, 200 µL of the incubated solution was mixed with 400 µL of denaturing buffer containing 0.2 M Tris-Cl (pH 6.8), 20% glycerol, 10% sodium dodecyl sulfate (SDS), 0.05% bromophenol blue and 10 mM β-mercaptoethanol and heated at 100°C for 10 min to stop the digestion. Proteolytic activity was determined on the Coomassie blue-stained gel after electrophoresis by observing the cleavage patterns of purified fibrinogen chains.

2.2.5.4. Inhibition of gelatinolytic activity

The effect of synthetic tripeptides and EDTA on purified protein was assayed using SDS-PAGE (4% stacking/10% resolving gel) to determine gelatin degradation. The purified protein (10 ng/µL) was incubated firstly with synthetic tripeptide (5 mM) or EDTA (100 µM) at 37 ˚C for 10 min. Then, 10 µL of gelatin solution (2 mg/mL in distilled water) was added and 20 µL of this incubated solution was taken out at 1 h and 20 h, mixed with 5 µL of 5x denaturing buffer and heated at 95˚C for 2 min. The cleavage patterns on gelatin produced by protease action were observed on Coomassie blue-stained gels after electrophoresis.

2.2.5.5. Inhibition of fibrinogenolytic activity

The effect of synthetic tripeptides or EDTA on purified metalloproteinase was assayed using SDS-PAGE (4% stacking/12% resolving gel) to determine fibrinogen degradation. The purified protein (32 ng/µL) was incubated firstly with synthetic tripeptide (5 mM) or EDTA (100 µM) at 37 ˚C for 10 min. Then, 10 µL of fibrinogen solution (2 mg/mL in distilled water) was added and 20 µL of this incubated solution was removed at 1 h and 20 h, mixed with 5 µL of 5x denaturing buffer (25% v/v Triton X-100 in dH_2O) and heated at 95˚C for 2 min. The cleavage effect on fibrinogen chains by the SVMPs were observed on Coomassie blue-stained gel following electrophoresis.

Chapter 3. Construction of a venom gland cDNA library

3.1. Introduction

The study of venom gland transcriptomes by identifying and cataloging toxin genes is helpful in predicting the presence of various toxin families in the venom and in understanding their pharmacological properties that may help predict their pathological symptoms. In addition the relative types and amounts of toxin genes in any venom gland provide useful information for characterization of isoforms of a toxin family for understanding the relationships among species.

Sequencing of snake venom transcriptome starts from cloning complementary DNA (cDNA) libraries constructed from snake venom glands. Cellular mRNA is extracted from the gland tissues, reversed transcribed to produce cDNA from the initial mRNA, and incorporated into a plasmid using appropriate vectors and adapters. The plasmids are then grown up in a suitable host such as *E. coli*. The cDNA clones contain a single DNA insert derived from a gene transcript (mRNA) and thus provide a library of the portions of the genome that are expressed in venom glands.

The ends of the cDNA insert are sequenced to generate expressed sequence tags (ESTs). Expressed sequence tags are single DNA sequencing reads made from automatic DNA sequencers. The ESTs are short, averaging approximately 400 base pairs (bp) in length. They generally represent only fragments of genes, not complete coding sequences. Each clone normally has a 5’ and 3’ EST associated with ends of the cDNA insert. ESTs are used for searching for new members of gene families in the same species (paralogs), for functionally equivalent genes in other species (orthologs), or for alternatively spliced forms of known genes (122).

A cDNA library comparison between neonate and adult *Bothrops jararaca* showed age and gender factors influence the variability of a snake venom gland transcriptome. The relative percentage of SVMP transcripts from neonate decreased from 53.2% to 29.9% in adult. This was accompanied by increases in other toxins such as BPP precursors, snaclecs and SVSPs. In comparison of adult male and female, SVMP transcripts were somewhat more abundant in the female venom gland, but no significant difference was found among P-classes of SVMPs between the male and

female libraries. The variability is also characterized by the structural diversity of SVMPs precursors found in newborn and adult transcriptomes (123). The analysis of the venom transcriptome of *Echis ocellatus* elucidated that SVMP transcripts are abundant and the most divergent containing all three (PI-PIII) SVMP transcripts classes. There are also several distinct isoforms of BPPs, PLA_2s, Snaclecs, SVSPs and SV-LAAOs (124).

In study of Sai-Ngam A (2007), a total of 135 expressed sequence tags (ESTs) was generated from the cDNA library of Thai Russell's viper venom glands. The most abundant toxin transcripts were phospholipase A_2 (13%), BPP-CNP precursor (3.7%), serine protease (3%), metalloproteinases (2.2%), nucleotidases (1.5%). Non-toxin ESTs such as dendritic cell protein (1.5%), tyrosine phosphatase (1.5%) and NADH dehydrogenase subunit 1 (1.5%) were identified in the cDNA library. The size of the library was approximately 1.0 x 10^6 plaque-forming units per microgram of DNA vector. The length of cDNA inserts varied from 0.6 kb to over 2.0 kb (average length = 0.9 kb).

3.2. Aims

Snake venom metalloproteinases are in a multi-gene family which has a broad spectrum of biological activities leading to severe bleeding and coagulopathy in snake bite patients. The MRV venom contains high proteolytic activity but SVMPs within it and its toxin profile are not yet known. The composition and complexity of proteins can be explored by transcriptomic analysis. Expressed sequence tags (ESTs) from a cDNA library of venom glands can provide a snapshot of active genes in the tissue, especially toxin genes that are highly expressed in venom glands. This approach can investigate the gene expression profile: the identification of transcripts, the abundance of transcripts. ESTs analysis has advantages such as 1) cDNA library construction is relatively fast and is inexpensive; 2) ESTs are mRNA containing only functional coding region; and 3) clones are short and easy to sequence.

The primary aim of this chapter is to generate expressed sequence tags (ESTs) of SVMPs transcripts from Myanmar Russell's viper venom glands and thus analyse its transcriptomic profile, especially the SVMP toxin family to detect its composition and

different isoforms and sequences to predict the biological activities responsible for clinical manifestations.

3.3. Results and discussion

3.3.1. Sample collection and mRNA isolation

The venom collected from individual snakes was dried separately in a dessicator under vacuum overnight until crystallization. The salivary glands of the vipers were dissected according to the protocol for collection of snake tissues for genetic study from the Snake Farm, Queen Saovabha Memorial Institute, the Thai Red Cross Society. The details of the sample collected are described in the following Table 3.1.

Table 3.1. Specification of snakes and their collected venom

Snakes	1	2	3	4
Sex	Male	Male	Female	Female
Weight (kg)	0.95	0.74	1.49	1.27
Length (inches)	48	46	49	47
Girth (inches)	6.0	4.5	7.0	6.0
Gland weights (g)				
Left	1.0	0.9	1.0	0.8
Right	1.0	0.8	1.0	0.8
Venom collected (mg)	47.5	121.5	203.6	83.4

From the 2 male snake venom glands, 1.4 mg of total RNA was extracted. Then, 14.7 µg of mRNA (in 2 ml volume) was isolated from the total RNA: this represents approximately 1% of the total RNA. The amount of mRNA after precipitation and concentration was 10.3 µg. The loss of mRNA may be due to washing during the isolation procedure.

3.3.2. cDNA library construction

Five µg of mRNA was used for cDNA synthesis. The chromatogram for the cDNA purification is shown in Figure 3.1. The total cDNA in E2 tube (2nd tube for collection of eluted sample) of the sample (475 ng) was approximately the same as that in E2 tube of 2.0 kb RNA control (575 ng) indicating that the purification procedure is comparable. There was more cDNA in the W1 tube in the eluted sample comparing to the eluted RNA control. It may probably be due to the subsequent small cDNA fraction was eluted together with residual adaptors and primers during washing step.

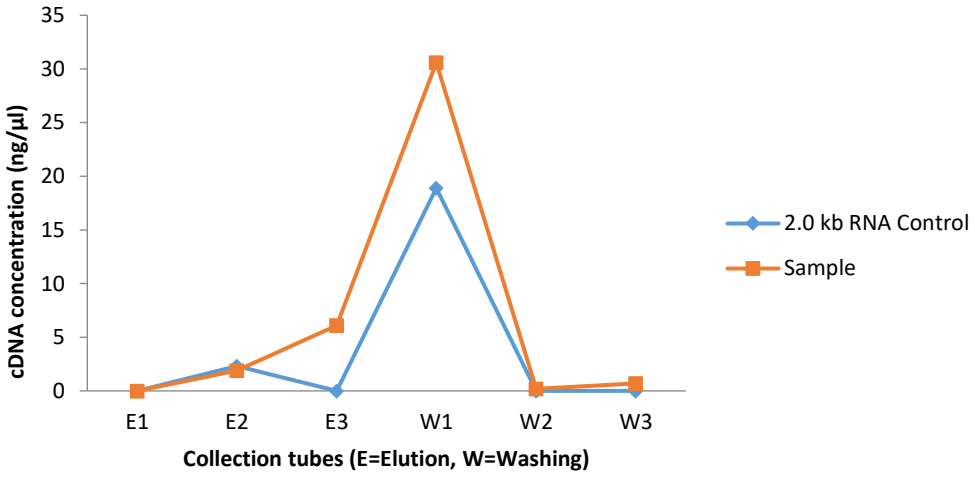


Figure 3.1. Chromatogram for cDNA size-fractionation through Sephacryl S-500 HR resin mini-column using TEN buffer. Three microfuge tubes: E1, E2 and E3 for collection of elute and another 3 microfuge tubes: W1, W2 and W3 to collect the washing flow-through.

3.3.2.1. Primary library

The first eluted cDNA from the sample and the 2.0 kb RNA control (i.e. Elution tube 2, E2) were precipitated with 100% ethanol. The yield of cDNA after ethanol precipitation is shown in the Table 3.2.

Table 3.2. The cDNA yield before and after ethanol precipitation for BP recombination reaction

Elution tube (E2)	Before ethanol precipitation	After ethanol precipitation	For BP recombination reaction
2.0 kb RNA Control	575 ng	84 ng	63 ng
Sample	475 ng	96 ng	72 ng

72 ng of cDNA were used for generation of a plasmid cDNA library. The primary library yield was 7.8 x 10^4 total colony forming units (cfu). The calculated cfu for controls and the primary library with reference levels are shown in the Table 3.3. The colonies found on LB agar plate with appropriate antibiotics to calculate the titre of the primary library, controls are shown in Figure 3.2.

Table 3.3. The calculated colony forming units of primary library construction

	BP negative control	BP positive control	pUC 19	2.0 kb RNA control	Sample
cfu	5.1 x 10^3 cfu/ml	5.1 x 10^6 cfu/ml	1.2 x 10^{10} cfu/µg DNA	3.7x 10^4 cfu/ml	6.5 x 10^3 cfu/ml (Total cfu= 7.8 x 10^4 cfu)
Reference	<0.3% of BP positive control (1.5 x 10^4 cfu/ml)	≥1 x 10^6 cfu/ml	≥1.0 x 10^{10} cfu/µg DNA	≥1 x 10^6 cfu/ml	5x10^6 –1x10^7 total cfu

The total cfu of the primary library is not represented well enough, probably owing to the low amount of cDNA available for BP recombination. The recommended starting quantity of cDNA required to produce a library containing 5-10 million clones is 75-100 ng.

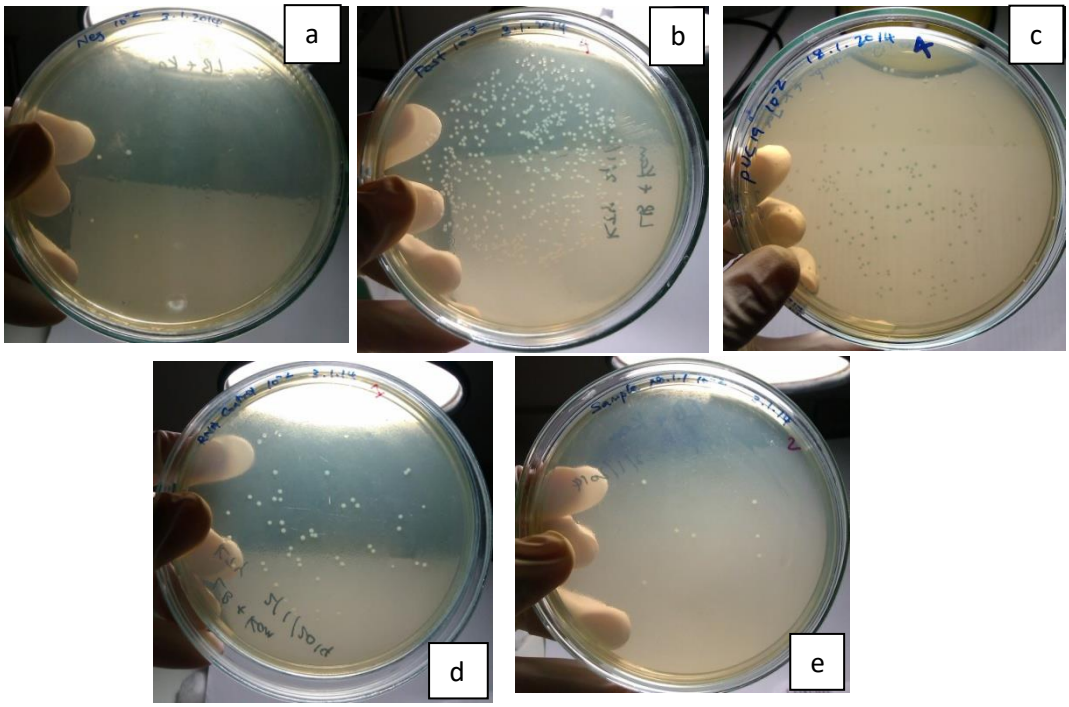


Figure 3.2. LB agar plate assay for primary cDNA library. a. Negative Control (10^{-2} dilution); b. Positive Control (10^{-3} dilution); c. pUC19 (10^{-2} dilution); d. RNA Control (10^{-2} dilution); e. Sample (10^{-2} dilution)

A total of 13 colonies were grown in 2 agar plates at 10^{-2} sample dilution. The purified plasmid from each colony was digested with *Bsr*G I restriction enzyme to predict the size of insert cDNA. The pDONR^{TM} 222 control shows a digestion pattern of 3 bands of 2.5 kb, 1.4 kb, and 790 bp. Each cDNA entry clone showed a vector backbone band of 2.5 kb and the specific insert bands. Figure 3.3 illustrates the digestion pattern of plasmid DNA and the insert size of individual clone is listed in the following Table 3.4.

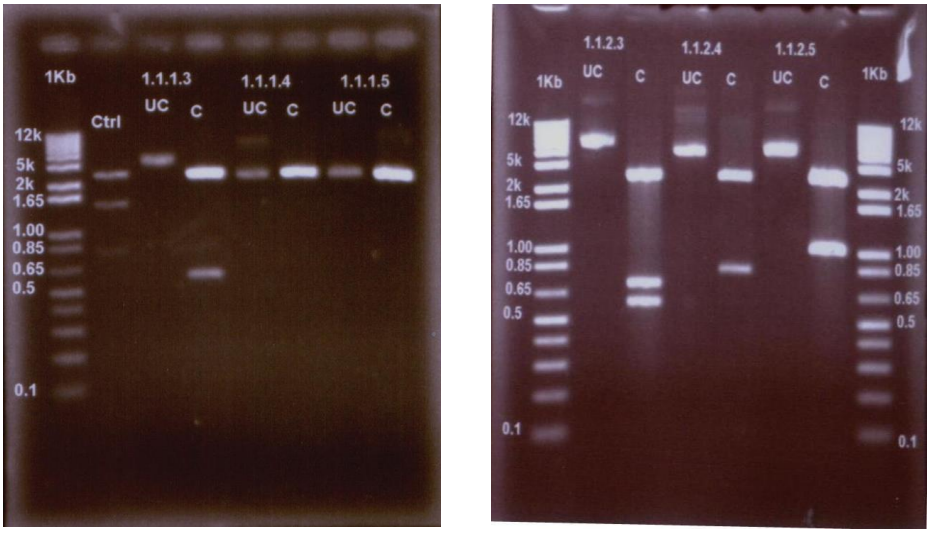


Figure 3.3. Agarose gel electrophoresis of *Bsr*G I digested cDNA plasmids. Ctrl = digested pDONR^{TM} 222 plasmid, UC= uncut plasmid of the sample, C= cut plasmid of the sample. A 100 bp Plus DNA ladder was used as the marker.

Table 3.4. The insert cDNA size in pDONR^{TM}222 plasmid of individual clones (10^{-2} dilution)

Clone	Band Size (kb)	Insert Size (kb)
1	0.40	0.40
2	0.45+0.3	0.75
3	0.65	0.65
4	No insert	-
5	No insert	-
6	0.45	0.45
7	0.50	0.50
8	No insert	-
9	0.6+0.75	1.35
10	0.85	0.85
11	1.00	1.00
12	No insert	-
13	No insert	-

Table 3.5. Percent recombination and insert size range of primary cDNA library

Number of clones analyzed	13
Number of clones containing inserts	8
Percent recombinants	62% [95%]
Average insert size (kb)	0.74 [≥1.5]
Insert size range (kb)	0.4 – 1.35

From Table 3.5, 8 of the 13 analysed clones contain inserts resulting in 62% recombinants with the average insert size of 0.6 kb representing the primary library. For standard cDNA libraries, an average insert size of ≥1.5 kb and at least 95% recombinants should have been obtained. The low yields may suggest that the cDNA going into the BP recombination reaction is either of poor quality or is insufficient in quantity. For the current primary library, it would probably be due an insufficient quantity of cDNA used for the BP recombination reaction.

3.3.2.2. Secondary library

Since the cDNA yield in tube E2 was less than 75 ng, a second cDNA library was generated from tube E3 from size fractionation. The yield of cDNA from tube E3 was 167 ng after precipitation and washing. The insert to vector ratio was 1:2.7 as the cDNA yield was > 100 ng. The reaction mixture for BP recombination was prepared as in Table 3.6.

Table 3.6. The amount of reagents used in the BP recombination for secondary library construction

	Sample	BP Controls	
		Negative Control	Positive Control
*att*B-flanked cDNA	7 µl	-	-
pDONR^{TM} 222(150 ng/ µl)	3 µl	1.6 µl	1.6 µl
pEXP7-tet (50ng/ µl)	-	-	0.5 µl
TE buffer, pH 8.0	4 µl	5.4 µl	4.9 µl
BP Clonase II enzyme mix	6 µl	3 µl	3 µl
	20 µl	10 µl	10 µl

The amount of cDNA used and ElectroMAX cells for transformation reaction for secondary library was shown in Table 3.7.

Table 3.7. The reaction volumes used in electroporation of cDNA into ElectroMAX cells

	pUC 19 Control (10 pg/µl)	BP Negative Control	BP Positive Control	RNA Control	Sample
Amount	1.0 µl	1.5 µl	1.5 µl	1.5 µl	1.5 µl
ElectroMAX cells	20 µl	20 µl	20 µl	20 µl	20 µl
Number of aliquots	1	2	2	2	6

The cfu calculated from plate assay for the second cDNA library were shown in Table 3.8.

Table 3.8. The calculated colony forming units of second cDNA library construction

	BP negative control	BP positive control	pUC 19	Sample
cfu	3.6 x 10^4 cfu/ml	1.4 x 10^6 cfu/ml	4.8 x 10^{10} cfu/µg DNA	2.0 x 10^3 cfu/ml (Total cfu = 2.8 x 10^4 cfu)
Reference	<0.3% of BP positive control (4.2 x 10^3 cfu/ml)	≥1 x 10^6 cfu/ml	≥1.0 x 10^{10} cfu/µg DNA	5x10^6 –1x10^7 total cfu

The second cDNA library was also not very representative (i.e., total cfu from sample cDNA library was less than 5 x 10^6) despite a higher starting quantity of over 100 ng. The colony count for the BP negative control is more than 0.3% of the BP positive control although the transformation efficiency is good (i.e., the calculated cfu from pUC 19 plates was more than 1.0 x 10^{10} per µg DNA). This might be due to the ethanol precipitation and washing before transformation causing the loss of cDNA used in the transformation reaction. Only 2 colonies were cultured from a 10^{-2} dilution sample plate.

To isolate more clones, a 10^{-1} dilution sample from primary library; undiluted and a 10^{-1} dilution sample from the secondary library were used to inoculate on LB agar plates containing kanamycin. From 2 libraries, 251 colonies were picked to determine the sizes of insert. Among them, 45 clones contained cDNA inserts (Table 3.9). The percentage of recombinants in primary library was 45.8% and this contains more plasmids with inserts than secondary library. The average insert size for the libraries was 0.86 kb.

Table 3.9. Characteristics of the cDNA libraries

	Primary library	Secondary library	Total number of clones cultured
Percent recombinants (95%)	45.8%	4.2%	
Average insert size (kb) (≥1.5kb)	0.86	0.63	
Insert size range (kb)	0.2-2.05	0.15-1.65	
Number of clones analyzed	83	168	251
Number of clones containing inserts	38	7	45

3.3.3. DNA sequencing and annotation of sequences

The cDNA sequencing was carried out using the Sanger method with both M13 Forward (-20) primer and M13 Reverse primer (5'-CAGGAAACAGCTATGAC-3') in separate PCR reactions. We used Finch TV to edit the DNA chromatogram and VecScreen to remove the vector sequence. BLAST programs were used for sequence annotation. From the sequencing results, 11 clones were toxin sequences, 12 clones contained toxin-related sequences and 19 clones showed cellular proteins sequences based on the significant similarity search against the NCBI database through Blastn. The clones which had no significant similarity against Blastn search (n=3) were annotated using their homologous sequences through a Blastx search (Table 3.10).

Table 3.10. Best matches of toxin and non-toxin encoding transcripts identified in conventional venom gland cDNA library construction

Groups	Genes
Toxin (n=11, 24.4%)	5'-nucleotidase, phospholipase A_2, C-type lectin-2, cysteine-rich secretory protein, Factor X activator light chain 2, C-type lectin and snake venom metalloproteinase
Toxin-related (n=12, 26.7%)	Snake venom metalloproteinase inhibitor, trypsin inhibitors
Non-toxin (n=19, 42.2%)	Acid ceramidase, multiple coagulation factor deficiency 2 (MCFD2), coiled-coil-helix-coiled-coil-helix domain containing 2(CHCHD2), cyclin F, microsatellite sequence, protein disulfide isomerase (PDI), solute carrier family 25 (peroxisomal membrane protein), focadhesin (FOCAD), 60s ribosomal protein L26-like, ribosomal protein S3a, P68 alpha subunit, ornithine decarboxylase 1 (DOC1), transmembrane protein 181 (TEMEM181), t-complex 1(TCP1), prolyl 4-hydroxylase beta polypeptide (P4HB), protein phosphatase 1 regulatory subunit 13B (PPP1R13B), mitochondria (complete gene), zinc transporter
No significant similarity sequences in Blastn search. (searched through Blastx) (n=3, 6.7%)	Regulator complex protein LAMTOR5 homolog, uncharacterized transcriptional regulatory protein YLR278C, ribosomal protein S6 kinase 2 beta

Table 3.10 lists the putative toxin protein identities from random collected clones from cDNA library of Myanmar Russell's viper venom gland. In the cDNA library of Myanmar Russell's viper venom gland, a total of 45 ESTs have been generated. The abundance of the toxin clones is as follow: phospholipase A_2 (3/11, 27.3 %), CRISP (3/11, 27.3%), C-type lectins (3/11, 27.3%), metalloproteinase (1/11, 9.1%) and

5’nucleotidase (1/11, 9.1%). From 11 toxin ESTs, PLA_2 (Clone 1.1.3.5) and cysteine-rich secretory protein (Clone 1.1.3.34, 1.1.4.12, 1.1.4.20) were found to have full-length sequence.

Table 3.11. A list of putative toxin protein identity matches for expressed sequence tags (ESTs) obtained from randomly sequenced clones from Myanmar Russell's viper venom gland

No.	Clone	Length (bp) of cDNA insert	BLAST Annotation	EST length & Bit score (E-value)	Species	Accession No.
1	1.1.2.3	813	5'-nucleotidase mRNA, partial cds	601 (0.0)	*Macrovipera lebetina*	KF408296.1
2	1.1.2.4	107	phospholipase A2-I mRNA, complete cds	80 (9.0×10^{-33})	*Daboia russellii russellii*	DQ365974.1
3	1.1.3.1	395	factor X activator light chain 2 mRNA, complete cds	242 (3.0×10^{-122})	*Macrovipera lebetina*	AY578116.1
4	1.1.3.5	710	phospholipase A2-II mRNA, complete cds	670 (0.0)	*Daboia russellii russellii*	DQ365975.1
5	1.1.3.34	960	cysteine-rich secretory protein Dr-CRPK mRNA, complete cds	927 (0.0)	*Daboia russellii*	EU589604.1
6	1.1.4.12	960	cysteine-rich secretory protein Dr-CRPK mRNA, complete cds	927 (0.0)	*Daboia russellii*	EU589604.1
7	1.1.4.13	597	factor X activator light chain 2 mRNA, complete cds	379 (0.0)	*Macrovipera lebetina*	AY578116.1
8	1.1.4.20	883	cysteine-rich secretory protein Dr-CRPK mRNA, complete cds	840 (0.0)	*Daboia russellii*	EU589604.1
9	1.1.4.29	242	C-type lectin-2 mRNA, complete cds	165 (1.0×10^{-79})	*Bitis gabonica*	AY429478.1
10	1.2.2.14	209	lebetase isoform Le-4 mRNA, partial cds	106 (7.0×10^{-47})	*Macrovipera lebetina*	AY987816.1
11	1.2.4.27	365	phospholipase A2-I mRNA, complete cds	501 (3.0×10^{-138})	*Daboia russellii russellii*	DQ365974.1

3.3.4. Analysis of full-length toxin sequences

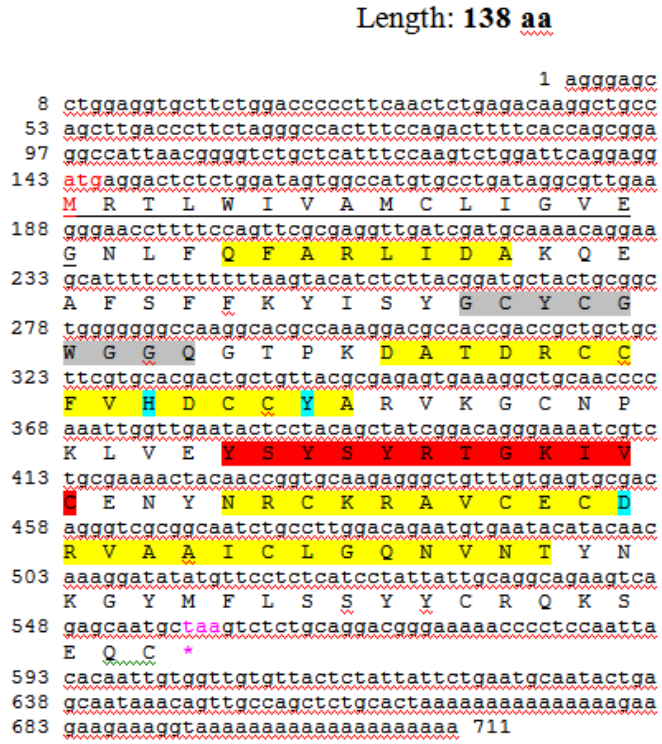
For sequence analysis, open reading frames were searched with ORF Finder and multiple sequence alignment was performed with ClustalW. The signal peptides were predicted by SignalP 3.0.

3.3.4.1. Phosholipase A_2

Figure 3.4 illustrates the deduced amino acid sequence of PLA_2. The signal peptide is underlined. Secondary structures such as α-helices (yellow), β-strands (red) and Ca-binding loops (grey) can be identified from the translated amino acid sequence. The three known catalytic amino acids, His48, Tyr52 and Asp99 are highlighted in blue-green (125).

There are 2 types of *D. russelli* venoms: type N and type S. Type N venoms, such as those of *D. russelli formosensis*, *D.r. siamensis*, and *D. r. russelli* contain PLA_2 with an N-terminal asparagine residue. Type S venom is seen in *D. russelli pulchella* and contains PLA_2 with an N-terminal serine residue (125).

On multiple sequence alignment, Myanmar Russell's viper PLA_2 (MRV PLA_2) has 84% identity to DbTx-B which was previously derived from the same species (Myanmar Russell's viper) (125), and 76% sequence identity to PLA_2 from Thailand (126) and Taiwan Russell's vipers (127) (Figure 3.5). Since DbTx-B and RV-4 have equivalent neutrotoxic



activity and are basic in nature, it can be predicted that our MRV PLA_2 protein might be a basic PLA_2 with similar neurotoxic function.

Figure 3.4. **The cDNA and deduced amino acid sequence of phospholipase A_2 from Myanmar Russell's viper.** The signal peptide is underlined, α-helices in yellow, β-strands in red, Ca-binding loops in grey, and 3 catalytic amino acids are in blue-green.

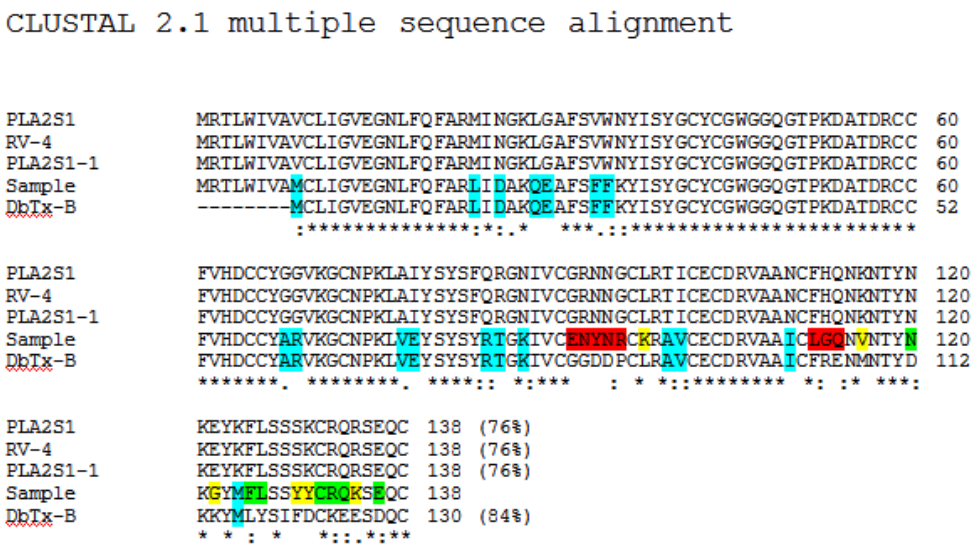


Figure 3.5. Comparison of the deduced amino acid sequence of PLA_2 (Myanmar Russell's viper) with those of other Russell's vipers. DbTx-B from *D. russelli siamensis* (Myanmar viper); RV-4 from *D.russelli formosensis* (Taiwan viper); PLA_2S1 and PLA_2S1-1 from *D. russelli siamensis* (Thailand viper). The same amino acid residue between Sample and DbTx-B are in blue-green. The different amino acid residues are in red, yellow and green.

3.3.4.2. Cysteine-rich secretory protein

The translated sequence for cysteine-rich secretory protein (CRISP) is shown in Figure 3.6. It shows the characteristic features of CRISP: 16 conserved cysteines (purple), 10 of which are clustered in the C-terminus. CRISPs from snake venom are single chain polypeptides (20-30 kDa) with two Zn^{2+} binding motifs (yellow and blue-green). They constitute a novel family of venom proteins and their functions are yet to be elucidated, though they appear to inhibit smooth muscle contraction (128).

The CRISP sequence from Myanmar Russell's viper (MRV CRISP) has 98% identity to Dr-CRPK from Taiwan Russell's viper (Gene Bank: EU589604.1) and 90% identity to sequences of CRISPs from European vipers (Figure 3.7). It should also be noted that the two Asian vipers' proteins possess an N-terminal serine residue whereas those of two of the European vipers have an N-terminal asparagine residue (129).

The finding of full-length CRISP is novel for Myanmar Russell's viper. Its exact biological functions on laboratory animals should be characterised in the future using venom-purified of recombinant proteins after cloning.

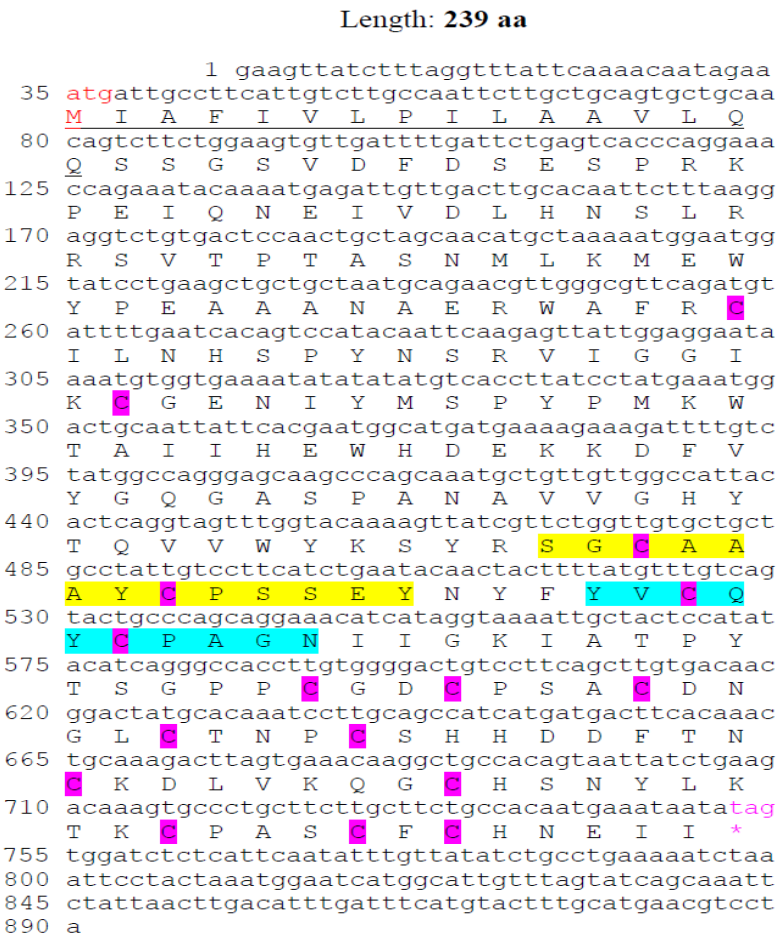


Figure 3.6. The cDNA and deduced amino acid sequence of cysteine-rich secretory protein (CRISP) from Myanmar Russell's viper. Two Zn^{2+} binding motifs: one in yellow and another in blue-green. Sixteen conserved cysteines are in purple.

Figure 3.7. Comparison of the deduced amino acid sequence of CRISP (Myanmar Russell's viper) with those of other vipers. Dr-CRPK from *D.russelli* (Taiwan Russell's viper). CRVP-VIPBE from *Vipera berus* (European viper); CRVP-VIPNI from *Vipera nikolskii* (Ukraine, Romania, Russia). Two Zn^{2+} binding motifs are underlined. The same amino acid residue between Sample and Dr-CRPK are in yellow. The different amino acid residues are in red, purple and green.

3.3.5. 5' Rapid amplification of the cDNA end (5'RACE) of partial-length toxin sequences

The clone (1.2.2.14) which gave the annotation of lebetase isoform Le-4 (a metalloproteinase from *Macrovipera lebetina*) possessed 209 bp sequence and was used for the design the Gene specific primers (GSPs) for 5'RACE. A 5'RACE reaction was performed to obtain full-length cDNA sequences. 5'RACE System for Rapid Amplification of cDNA Ends kit (Invitrogen, Life technologies, Carlsbad, California) was used. GSP1 was used for synthesis of the 1st strand cDNA. The mRNA was then degraded with RNase. The

cDNAs were purified through a SNAP column. Then, the purified cDNA was tailed with oligo-dC and the tailed cDNA later amplified with GSP2 and Abridged Anchor Primer.

Primer	Sequence	nt	Tm	CG
GSP1 (Yellow)	5'-ACTTGTTGGCCAAAGGC-3'	17	52°C	53%
GSP2 (Red)	5'-ATGTACAACACGAAGGGCTTATTTGC-3'	26	74°C	42%

The PCR conditions for dC-tailed cDNA products was as follows: initial denaturation for 5 mins at 94°C, followed by 35 extension cycles (denaturation for 1 min at 94°C, annealing for 1 min at 64°C, and elongation for 2.5 min at 72°C), and a terminal extension step of 72°C for 10 min.

In Figure 3.8, an 800 bp PCR product was obtained from the metalloproteinase clone. This clone showed the annotation of disintegrin on Blastn search after DNA sequencing of the 5'RACE product. This is a mismatch with the annotation result obtained from partial EST sequencing due to the fact that NCBI classification of metalloproteinase and disintegrin are in one group. Thus, it is difficult to get true annotation with partial coding sequencing especially for multi-domain protein classes.

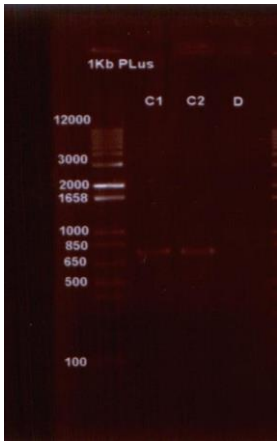


Figure 3.8. 5'RACE products of metalloproteinase clone. Lane 1 (Molecular marker); Lane 2 (C1): cDNA (50ng) with dC-tail; Lane 3 (C2): cDNA (100ng) with dC-tail; Lane 4 (D): cDNA without dC-tail

Length: **110 aa**

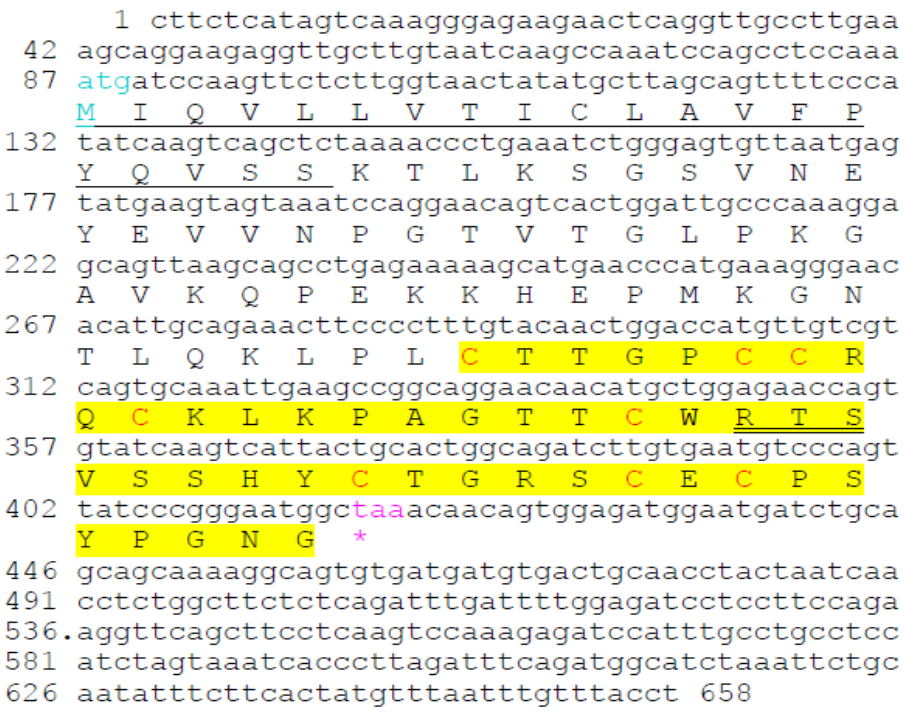


Figure 3.9. The cDNA and deduced amino acid sequence of disintegrin from Myanmar Russell's viper. The signal peptide is underlined. The active motif which binds to its target integrin is double-underlined. The disintegrin domain in yellow. The cysteine residues are in red letters.

In Figure 3.9, the signal peptide of the translated amino acid sequence of disintegrin is indicated by single underline. The active tripeptide RTS is double-underlined. The 8 cysteine residues are highlighted in red. There are 5 groups of snake venom disintegrin according to their polypeptide length and numbers of disulphide bonds: 1) Short integrins are composed of 41-51 residues and four disulfide bonds; 2) Medium-sized disintegrins are about 70 amino acids and have six disulfide bonds; 3) Long integrins are with an ~ 84-residue polypeptide cross-linked by seven disulfide bonds; 4) PIII-disintegrins contain an N-terminal disintegrin-like domain of about 100 amino acids including 16 cysteine residues involved in the formation of eight disulfide bonds, and a C-terminal 110-120-residue domain crosslinked by six disulfides; 5) Dimeric disintegrins

contain subunits of about 67 residues with 10 cysteines involved in the formation of four intra-chain disulfide bonds and two inter-chain cysteine linkages (85). Our MRV disintegrin sequence belongs to a class of short-sized disintegrins because it contains 8 cysteine residues and is 44 amino acids in length. It is 99% identical to disintegrin from *Trimeresurus jerdonii*. It has an active sequence RTS instead of conventional RGD motif for integrin binding. The novel RTS-disintegrin from *Trimeresurus jerdonii* inhibits $\alpha_1\beta_1$ integrin (130) which is a very selective receptor of basement membrane type IV collagen and laminin-1 and block the adhesive properties of microvascular endothelial cells (131). This could lead to extravasation of erythrocytes through disturbed-junctions between endothelial cells and the basement membrane (132).

3.4. Conclusions

The cDNA library of Myanmar Russell's viper venom glands contained only 11 toxin ESTs out of total 45 ESTs: the library was thus a low representative of the venom transcriptome. Only one expected metalloproteinase clone was found in the current library, of partial length, and it is actually a disintegrin sequence after amplification by the 5'RACE reaction. However, the finding of full-length PLA_2 and CRISP transcripts gives reliable structural-functional information from their deduced amino acid sequence. Thus, because the library was small and only contained a relatively few clones, the abundance and different types of SVMP transcripts could not be accurately quantified.

This is one of the major disadvantages of EST analysis by this approach of cDNA cloning. All clones from the constructed cDNA library are unlikely to be full-length when they are oligo dT primed. This is particularly true for large transcripts. Thus, it would be difficult to obtain the full-length of long SVMP transcripts. Some clones will be sequenced more than once, potentially duplicating EST reads. Thus, only a portion of transcripts is analyzed and isoforms are generally indistinguishable from each other. When the library is this small, then low abundance transcripts are unlikely to be cloned and a further limitation is that it is very difficult to quantify relative transcript abundance.

The studies of snake venom-gland transcriptomes with cloning of cDNA libraries and Sanger sequencing yield a low coverage and a high likelihood of missing and partially sequenced transcripts. This might be due to 1) the potential of the mRNA transcripts in plasmids to be partially expressed in their bacterial cells with lethal effects and 2) small cDNA fragments are over represented compared to larger ones, due to a higher transformation efficiency of smaller plasmids (133).

Chapter 4. Next-generation sequencing of snake venom glands and *de novo* assembly of transcriptome

4.1. Introduction

Next-generation sequencing (NGS) has advantages over Sanger sequencing as follows.

1) *In vitro* construction of a sequencing library, followed by *in vitro* clonal amplification generating sequencing features to overcome the restriction of parallelism of conventional sequencing.
2) Lower costs for DNA sequence production because of decreasing the effective reagent volume per feature in the scale of picolitres or femtolitres.
3) RNA-sequencing by NGS is non-biased and provides a profile of the entire transcriptome of a cell or tissue.
4) Quantitation, provided there is adequate sequence coverage, is quantitative and high and low abundance transcripts can be identified and quantified.

In contrast, these advantages are offset by the following factors.

1) The read-lengths are much shorter than conventional sequencing. The read-length varies from 13-250 bp according to the platforms used for NGS. Sanger sequencing can achieve read-lengths of up to ~ 1,000 bp. In some exceptional NGS platforms, Heliscope, PacBio of MinION, sequencing of full transcripts up to 10 kb in length is possible.
2) The per-base 'raw' accuracy in Sanger sequencing is as high as 99.999% whereas the raw accuracy of base-calls generated by NGS platforms are generally tenfold less accurate than base-calls generated by Sanger sequencing but this can be corrected by multiple reads.
3) Mapping of reads is easier when the genome of the species has already been sequenced, but this is not always the case.

The NGS provides millions of short reads at a much cheaper cost per bp, whereas the Sanger sequencing provides longer reads for bridging, but at a much higher price per bp.

Having both pros and cons, NGS is preferred for large-scale projects due to the reduction in the per-base cost of sequencing and its high-throughput nature. The principal challenge when using NGS is downstream data management as follows (134).

The sequence reads obtained from the NGS platforms must be assembled to generate the full-length transcripts. Depending on whether a reference genome assembly is available, transcriptome assembly strategies fall into one of three categories: a reference-based strategy, a *de novo* strategy or a combined strategy that merges the two.

De novo assembly was chosen to use for our sample data set as there is no reference genome for the Russell's viper species. *De novo* assembly can be performed to obtain the correct alignment of reads without a reference genome and knowledge of known splice sites.

Most *de novo* assemblers such as Trinity, Trans-ABySS, Oases and CLC Genomics WorkBench (*de novo* assembly) use a de Bruijn graph-based approach to reconstruct the transcripts from the data set of reads (135).

4.2. Aims

The transcriptome analysis using ESTs derived from cDNA library in the previous chapter was limited and only detected the SVMP toxins family, due to low coverage and generally was not quantitative. The SVMPs are proteins of a multigene family with highly diverse isoforms. Thus, a better technique with deep-sequencing is needed to observe a complete picture of expression profile of SVMPs.

The development of high-throughput, less expensive next-generation sequencing (NGS) should alleviate issue of low coverage and provide a more comprehensive characterisation of the genes contributing to snake venoms. Thus, to get more support for the qualitative and quantitative analysis of SVMPs transcripts, RNA-Sequencing of Myanmar Russell's viper was then performed using NGS.

The primary aim of this chapter was to generate SVMPs sequences from Myanmar Russell's viper venom glands in good coverage and to analyse their transcriptomic profile.

4.3. Results and discussion

4.3.1. RNA-Sequencing of snake venom gland

There are two limitations in *de novo* transcriptome assembly as follows.

1) *De novo* transcriptome assembly requires a much higher sequencing depth for full-length transcript assembly than does the reference-based assembly strategy.
2) *De novo* transcriptome assemblers are very sensitive to sequencing errors and the presence of chimeric molecules in the data set (135).

To overcome the first limitation, the mRNA was enriched from total RNA in our experiment. The aim of this was to increase the sequencing depth since mRNA accounts only for 1-4% of the whole transcriptome. In addition, the HiSeq 2000 platform can give sequencing coverage c=30x (>10^9 reads per run) at 100 bases read length for human genome or exome sequencing (136).

To cope with the second limitation, the sequencing error was reduced by using Trimmomatic and validated the data set with Phred score (Q20), GC content (%) before assembly.

The purity of mRNA from male and female snake samples was measured in a NanoDrop spectrophotometer. The 260/280 ratios of samples were 1.96 and 2.17 respectively (reference: a ratio of ~2.0). The Truseq RNA library QC result showed 105 ng/µL (574 nM) concentration and 281 bp size for the male snake library and 99 ng/µL (540 nM) concentration and 284 bp size for the female snake library.

From Illumina HiSeq2000 platform, 82,251,232 reads from the male library and 79,578,046 reads from the female library were generated. After removal of low-qulaity reads and adapters using Trimmomatric, 78,647,670 reads with 98.93% Q20 bases (base

quality more than 20 and an error rate of less than 0.01) from the male library and 74,444,200 reads with 98.55% Q20 bases from the female library remained for *de novo* assembly.

4.3.2. *De novo* assembly of transcriptome

For *de novo* transcriptome assembly, 2 diverse approaches were performed using Trinity (r20140717) and CLC Genomics WorkBench. (*de novo* assembly). Both use a de Bruijn graph-based approach to reconstruct the transcripts from the data set of reads with different k-mer values: 25 in Trinity and 45 in CLC. Then, their assembly performance was assessed with QUAST.

Comparing transcriptome assemblies from Trinity and CLC assemblers (Table 4.1), the percentage of total useable reads was 82% in the samples. The GC content (%) was approximately 44% in corresponding samples. However, the total number of bases in the assemblies of 2 samples (Total length^) assembled by Trinity was higher than those assembled by CLC respectively. Similarly, N50 and total contig values were higher in assemblies of Trinity than in those of CLC. In addition, the length of the largest contig obtained by Trinity was longer than that obtained by CLC. Mismatches were seen (96 - 144 mismatches per 100,000 aligned bases) in assembles of CLC, while there was no mismatch in assemblies of Trinity. The quality assessment showed that Trinity had a better performance in transcriptome assembly than CLC on out data set, thus Trinity assembly data set was used for further analysis.

Table 4.1. QUAST quality assessment report for two assemblers

Parameters for quality assessment	Trinity Assembly		CLC Genomics Workbench (*De novo* Assembly)	
	Male	Female	Male	Female
# proper assembled contigs	88523	50858	59578	38432
# misassembled contigs	19326	10673	12553	8378
Total used reads (%)	82.04	82.65	82.59	82.10
Total length ^(Mb)	75	39	51	31
N50 (bp)	2516	1935	2446	1814
Total contig*	33106	20248	23791	16930
Largest contig (bp)	31422	17302	29849	15013
GC content (%)	43.87	44.66	43.72	44.62
# N's per 100 kpb#	0.00	0.00	144.97	96.38

^ The total number of bases in the assembly.

N50 = the length for which the collection of all contigs of that length or longer covers at least half an assembly.

*The total number of contigs of length ≥500 bp

#The average number of uncalled bases (N's) per 100,000 assembly bases. (i.e, mismatches)

4.3.3. Annotation and transcript abundance analysis of toxin transcripts

Contig gene name annotation was archived through Blastx (Gene Ontology) against protein sequence database (go_v20140820) and Blastn against collected serpent nucleotides database (142,183 sequences). The Blastx search revealed that 30,090.8 (34.06%) for male and 214,159.8 (47.50%) for female snake sequences have a significant blast hit. The Blastn search showed that 23,525 out of 88,325 transcripts (26.63%) in the male sample and 20,355 out of 50,858 transcripts (40.02%) in the female sample have a best hit after filtering.

The toxin groups were grouped manually based on the key words of individual toxin names. The FPKM values of contigs within a toxin group were combined to represent the individual toxin group. In general, the most abundant venom toxin family in both male and female samples are metalloproteinases (MPs), C-type lectins (CTLs), phospholipase A_2 (PLA_2s) and serine proteases (SPs). These contribute the major local and systemic effects seen in Russell's viper bite patients: potent haemorrhagic activity by MPs; coagulopathy by CTLs, SPs and disintegrins; myonecrosis by MPs and PLA_2s; inflammatory responses by MPs, SPs, PLA_2s; and cardiovascular actions by MPs, PLA_2s and SPs (48). There is considerable sex-specific variation in the content of the major toxin. The proportion of MPs and CTLs in the male is similar to the female, while SPs are more abundance in the female. This could predict that more severe coagulopathy is likely to occur clinically for envenoming by the female compared to the male. Among the major toxin families, the metalloproteinase transcripts were found to be the highest expressed among other toxins in both male and female snakes (Table 4.2, Figure 4.1).

4.4. Conclusions

This is the first transcriptome analysis in Russell's viper species and will provide rich information for structural-functional studies, new toxin discovery, antivenom development and evolution studies of Russell's viper. From this preliminary study, we have found that there may be gender-related differences in relative abundances of major toxin transcripts. The metalloproteinase transcripts are the most highly expressed in both male and female venom glands. However, the second most abundant toxin transcripts are C-type lectins in male, and serine proteases in female. The third highly expressed toxin group is phospholipase A_2 in male and C-type lectin in female. The differential abundance in toxin transcripts are closely associated with the biochemical and biological properties for adult male and female venoms (123). The sex-based difference in venom composition might be related to their diet. In *Bothrops jararaca* species, there is sex-based variation in biological properties of venom (137), and sexual divergence in diet (females feed primarily on larger conger eels whereas males feed upon smaller moray eels) in the Fujian sea snakes, *Laticauda colubrine* (138).

In previous ESTs library of Russell's viper venom gland, only few toxin genes could be identified and the clone cannot contain long genes. In RNA-Seq library, much more toxin genes can be detected and the whole transcripts can be assembled by Trinity.

Therefore, Next-Generation sequencing provided a more comprehensive coverage for transcriptomic characterizations of venom glands of snakes both in quantitative and qualitative ways than low-throughput sequencing approaches such as cloning and ESTs generation. RNA-Seq is more efficient and needs less time than EST sequencing for investigation of a transcriptomic profile.

Table 4.2. Relative abundance of the toxin groups between male and female Myanmar Russell's viper venom glands

Toxin group	FPKM value	
	Male	Female
Metalloproteinase	83525.39	32241.15
C-type lectin	82775.35	19769.36
Phospholipase A2	26114.33	8379.33
Serine protease	20869.60	20051.87
VEGF	15317.52	4270.38
L-amino acid oxidase	6262.55	1895.32
Cysteine-rich secretory protein	3575.78	800.20
Nucleotidase	1479.16	444.19
Nerve growth factor	1068.99	660.3
Phosphodiesterase	928.72	266.44
Aminopeptidase	295.17	150.44
Dipeptidyl peptidase	245.06	44.28
Hyaluronidase	139.67	33.13
Cystatin	129.52	37.10
Phospholipase B	127.62	62.10
Carboxypeptidase	103.60	23.06
3-Finger toxin	92.65	167.57
Venom factor	79.47	5.38
Matrix metalloproteinase	79.11	25.44
C-type naturetic peptide	52.47	0.94
Cardiotoxin	51.80	11.54
Phospholipase D	44.02	3.21
Vesp	35.58	16.90
Acetylcholinesterase	23.86	3.20
Waprin	15.81	1.56
Acid phosphomonoesterase	8.29	3.68
Phospholipase C	5.88	0
Myotoxin	5.87	0
Paraoxonase	0.74	0

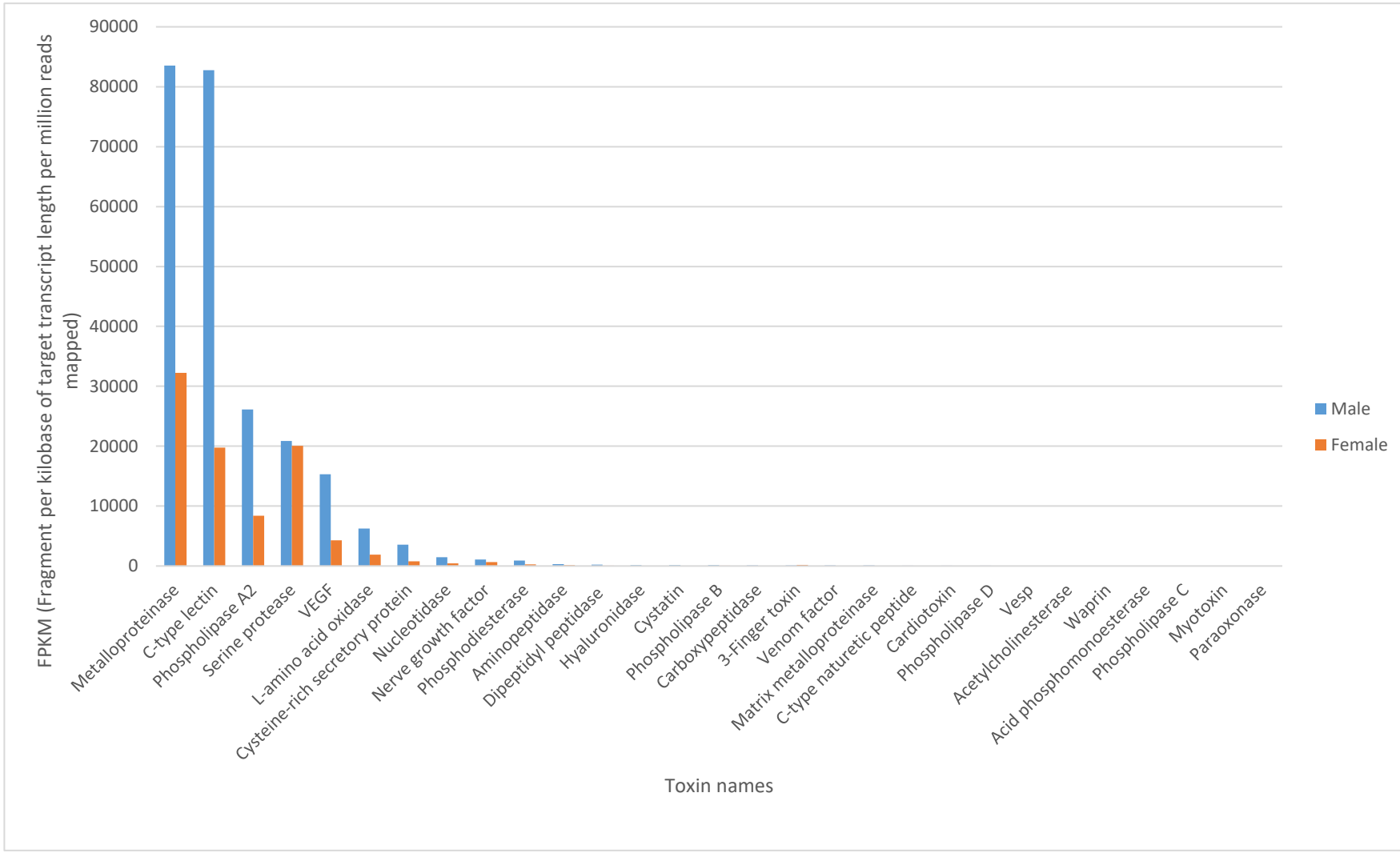


Figure 4.1. Relative abundance of the toxin groups in Myanmar Russell's viper venom glands between male and female sample in term of FPKM value.

Chapter 5. Analysis of snake venom metalloproteinase and SVMP inhibitor transcripts

5.1. Introduction

Transcriptome analyses are useful for 1) cataloging toxin genes expressed in the venom gland and 2) identification of novel toxins in existing families, as well as new families of toxins. The complexity of snake venom within a single species can be extremely high. Snake venom metalloproteinases (SVMPs) are the major toxin component of viper venoms and the SVMP family is one of the most abundant protein families in their venoms. It composed of protein members with variable structural motifs and functional activities. The presence and structural features of the domains characterize the function and toxicity of SVMPs. The diversity of SVMPs evolved by gene duplications followed by positive selection of non-synonymous mutations accumulated in genomes under strong positive adaptive selection. Sanz and Clavete (2016) suggested that the evolutionary history of SVMPs occurred via insertions and deletions of intronic regions to form the multi-locus SVMP gene multifamily. In addition, repetitive intronic elements, such as the long interspersed elements (LINEs) and the short interspersed elements (SINEs), occurred in several introns of SVMP genes. These interspersed repeats can influence gene expression (139). The insertion of interspersed repeats into a new genomic position may introduce promoters or enhancer sequence motifs for transcription of nearby genes and alternative splicing sites or polyadenylation sites, resulting in a change of overall level of gene expression (140, 141).

The differential expression of toxin groups, according to snake geographical distribution, sex or ontogenesis, has been previously described. The venomic study on *Bothrops asper* (Lancehead pitviper) showed that there was an ontogenic shift from a PIII-SVMP-rich to a PI-SVMP-rich compositions (142). A proteomic study among *Bothrops jararaca* siblings (7 male and 11 female siblings) showed the presence of sex-based variation in venom components especially in its P-III classes. Female venoms showed gelatinolytic activity associated with the P-III and P-I classes but it was clearly absent in the male venoms of

the *B. jararaca* siblings. The female venoms also showed the presence of more active caseinolytic enzymes than the male venoms (137).

5.2. Aims

This Chapter has 2 main aims: 1) to study the expression profile of SVMPs and 2) to study the snake venom metalloproteinase inhibitor (SVMPI) sequences from Myanmar Russell's viper venom glands.

Analysis of the expression profile of SVMPs from Myanmar Russell's viper venom glands could predict their structural-functional activities which contribute the significant clinical features of patients with envenomation. In addition, comparison with SVMP sequences from other Russell's vipers could give insights into the reasons for different signs and symptoms among Russell's vipers bites from different countries.

Thus, the objectives for first aim are to 1) identify, quantify, characterise and compare SVMPs transcripts from male and female Myanmar Russell's vipers; and 2) to compare the SVMP toxin sequences profile of Myanmar Russell's viper with those of other vipers.

For the second aim, endogenous tripeptides, called SVMPIs, of great interest for their attenuating effect on SVMPs because of the limited efficiency of antivenom treatment. SVMPI transcripts, encoding tripeptide inhibitors, were found to be expressed in African vipers and these peptide inhibitors in the venom were known to inhibit the SVMPs. However, SVMPIs from Myanmar Russell's viper and their effect towards SVMPs are still largely undefined.

Thus, the objective for the second aim is to 1) identify SVMPI transcripts from Myanmar Russell's viper venom gland transcriptome, and 2) to compare and analyse SVMPI transcripts from Myanmar Russell's viper with other vipers.

5.3. Results and discussion

5.3.1. Relative abundance of snake venom metalloproteinase transcripts

The contigs with metalloproteinase annotation (n=109, Male = 69 & Female = 40) were grouped according to their gi numbers obtained from Blastn search against Serpents nucleotide sequences from NCBI database. Subsequently, 59 contigs (Male=37 & F=22) with SVMP annotations were clustered into 4 disintegrin, 7 P-III and 2 P-II SVMPs-different clusters (Table 5.1). The contigs from each cluster, according to a strong homology between their overlapping regions, are likely to have originated from the same gene. The presence of 3'-untranslated regions of toxin mRNA points to incompletely sequenced transcripts. Some transcripts contained parts of both exon and intron indicating that they were still in pre-mRNA stage at the time of venom gland extraction or that there were potentially new splice variants for the transcripts. Interestingly, the presence of a transposable element, SINE was seen in the SVMP transcripts, that may play role in appearance of differently spliced pre-mRNAs (139).

For detailed structural analysis, contigs containing full-length and partial length CDS were then translated and compared to the translated sequences/ amino acid sequences from the Blastn annotation results. The translated sequences that were not matched with Blastn annotation result were also compared to the Blastx annotation results. After Blastx hits against UniProtKB/Swiss-Prot (swissprot) database, 2 contigs from the P-III 2 cluster from female sample matched to 2 different reference contigs c11797_g1_i1_F (disintegrin, full-length CDS) and c11797_g1_i4_F (VLAIP-A, partial length CDS). Similarly, out of 6 contigs containing partial length CDS from the P-III 7 cluster from male sample showed 2 different reference transcript matches: c9453_g1_i1_M, c66080_g1_i1_M (homologous to *Protobothrops mucrosquamatus* zinc metalloproteinase- disintegrin-like NaMP) and; c22734_g1_i1_M, c34552_g1_i1_M, c41009_g1_i1_M, c71069_g1_i1_M (homologous to *Echis coloratus* svmp-u mRNA) (Table 5.2).

As shown on Table 5.3, the abundant distribution of SVMP toxin transcripts: disintegrin (75%), P-III SVMPs (25%) and P-II SVMPs (approximately 0.002 %), were the same for

male and female samples indicating that the relative proportion of toxin transcripts is not directly related to the size of the cDNA library. In another way, it could be observed that the highly expressed contigs from both viper sexes shared the same identity: P-III SVMPs were found to be more highly expressed than P-II in Myanmar Russell's viper venom glands. No P-I SVMP transcripts were detected by our analysis. There are no other transcriptome data for *Daboia* genus to compare our findings to. However, the relative abundance of different classes of SVMPs shown by our study is similar to that of urutu (*Bothrops alternatus*) (143) from the same family *Viperidae*. The abundance of P-III class SVMP transcripts (assuming that transcripts are proportionally translated) might explain the intense pro-coagulant activity of *Viperidae* snake venom, and high proportion of disintegrin transcripts may correspond to a characteristic endothelial damage-specific for MRV venom toxicity.

The comparison of content of contigs encoding toxins in adult male and female venom glands showed some gender-related differences. A disintegrin transcript isoform (Dis 1b) was highly expressed only in the female venom gland. Some P-III SVMP isoforms (P-III 6, 7a and 7b) were only expressed in the male venom gland but at low expression levels. The P-II SVMP transcripts expressed in male and female were found to be different gene isoforms. This differential abundance in SVMP transcripts provided they are proportionally translated would be of casual significant for the biochemical and biological properties of the two analysed individual snakes and may reflect differences between adult male and female MRV venoms.

Among 3 classes, P-III SVMPs showed more diverse isoform contents. This extensive P-III SVMP gene diversification may reflect a presence of particular strong evolutionary pressure towards increase in functional diversity of SVMPs. This might have a significance in adaptation to different types of prey. The transcriptomic analysis of *Bothrops jararaca* revealed that stronger coagulant activity of newborn snake venom is strictly related to its SVMP quantitative variability (higher content of SVMPs than in adult venom glands). This variability is also characterized by the structural diversity of SVMP precursors found in new born and adult transcriptomes (123). There are also differences in substrate

specificity among new born and adult P-III class SVMPs. The distinct SVMP substrate specificities in venoms could be related to the structural diversity varied among SVMP precursors in venom glands. Most of the newborn P-III class SVMPs tend to play procoagulant activity, acting upon FII/FX or both (144). This ontogenic shift in *Bothrops jararaca* was explained by type of prey they consumed: feeding on ectothermic prey (mainly arthropods, lizards, and amphibians) through the juvenile phase and on endothermic animals (mainly small mammals) during adult life. Thus, prey death/immobilization is the main function of the venom of juvenile snakes and prey digestion is the main function for adult venoms (145).

The cDNA libraries comparison between neonate and adult *Bothrops jararaca* showed that age and gender factors influenced the variations in snake venom gland transcriptomes. The relative percentage of SVMP transcripts reduced from 53.2% in neonates to 29.9% in adults. This was accompanied by increase in number of transcripts for other toxins such as bradykinin-potentiating peptide (BPP) precursors, snake C-type lectin-like proteins (snaclecs) and snake venom serine proteases (SVSPs). SVMP transcripts were more abundant in the female venom gland, but no significant difference was found for P-classes of SVMPs between the male and female cDNA libraries. The variation is also characterized by the structural diversity of SVMP precursors found in newborn and adult transcriptomes (123).

The analysis of the venom transcriptome of *Echis ocellatus* illustrates that SVMP transcripts are abundant with the most divergent ones containing all three (PI-PIII) SVMP transcripts classes. There are also several distinct isoforms of BPPs, phospholipases A_2 (PLA_2), Snaclecs, SVSPs and snake venom L-amino acid oxidases (SV-LAAOs) (124).

Table 5.1. List of SVMP transcripts with their expression level in FPKM value and portion of mRNA

Classes	Annotation	Male		Female	
		FPKM	Portion of mRNA	FPKM	Portion of mRNA
Dis 1	gi|66390955|gb|AY987816.1| *Macrovipera lebetina lebetase* isoform Le-4 mRNA partial cds	62514.52	Full CDS	18456.39	Full CDS
Dis 2	gi|95007580|emb|AM261811.1| *Macrovipera lebetina transmediterranea* ml-G1 gene for VGD-containing dimeric disintegrin subunit ML-G1 precursor exons 1-2	82.5	intron	10.84	intron
Dis 3	gi|110346543|emb|AM286800.1| *Echis ocellatus* gene for MLD-containing dimeric disintegrin subunit exons 1-2	62.42	intron	6.93	intron
Dis 4	gi|95007582|emb|AM261812.1|*Macrovipera lebetina transmediterranea* ml-G2 gene for MLD-containing dimeric disintegrin subunit ML-G2 precursor, exons 1-2	2.46	intron	-	-
P-III 1	gi|300079899|gb|GQ420354.1| *Daboia russellii russellii* factor X activator heavy chain mRNA complete cds	9225.14	Full CDS	3829.48	Full
P-III 2	gi|61104774|gb|AY835996.1| *Macrovipera lebetina* VLAIP-A mRNA complete cds	5794.55	Partial CDS	7846.96	Full, Partial CDS
P-III 3	gi|727360728|gb|GBUG01000035.1| TSA: *Echis coloratus* svmp-n mRNA sequence	5490.9	3'UTR	2012.55	3'UTR
P-III 4	gi|83523625|emb|AM039691.1| *Echis ocellatus* mRNA for Group III snake venom metalloproteinase (Svmp3-Eoc1 gene) clone Eo_venom_04E07	139.71	Partial CDS, intron, 3'UTR, SINE	25.76	Partial CDS, Intron, 3'UTR, SINE
P-III 5	gi|387014235|gb|JU173711.1| TSA: *Crotalus adamanteus* Cadam_SVMPIII-2e mRNA sequence	5.66	3'UTR	3.34	3'UTR
P-III 6	gi|387014223|gb|JU173705.1|TSA: *Crotalus adamanteus* Cadam_SVMPIII-1a mRNA sequence	14.51	3'UTR	-	-
P-III 7	gi|727360714|gb|GBUG01000042.1| TSA: *Echis coloratus* svmp-u mRNA sequence	6.43	Partial CDS, 3'UTR	-	-
P-II 1	gi|320579332|gb|GU594194.1| *Echis ocellatus* clone 04C07 group II snake venom metalloproteinase (Eoc00006) mRNA partial cds	1.41	Partial CDS	-	-
P-II 2	gi|31322300|gb|AY204244.1| *Gloydius saxatilis* metalloproteinase/disintegrin saxin precursor mRNA partial cds	-	-	8.44	Partial CDS

CDS = coding DNA sequence

Table 5.2. Annotations of SVMP contigs containing full-length and partial length CDS against Swiss-Prot NCBI database

Classes	Sub-classes	Contigs	Annotation	Coverage	% identity	e-value
Dis 1	Dis 1a	c13890_g2_i1_M	gi|48428040|sp|Q7ZZM2.1|Disintegrin jerdostatin (*Protobothrops jerdonii*)	100%	100%	5e-81
		c11797_g1_i2_F	gi|48428040|sp|Q7ZZM2.1|Disintegrin jerdostatin (*Protobothrops jerdonii*)	100%	100%	5e-81
P-III 1		c20464_g1_i1_M	ADJ67475.1 factor X activator heavy chain [*Daboia russellii russellii*]	100%	97%	0.0
		c11797_g1_i3_F	ADJ67475.1 factor X activator heavy chain [*Daboia russellii russellii*]	100%	97%	0.0
P-III 2		c20464_g1_i2_M	gi|82228619|sp|Q4VM08.1|VM3VA_MACLB| Zinc metalloproteinase-disintegrin-like VLAIP-A (*Macrovipera lebetina*)	82%	89%	0.0
		c11797_g1_i4_F	gi|82228619|sp|Q4VM08.1|VM3VA_MACLB| Zinc metalloproteinase-disintegrin-like VLAIP-A (*Macrovipera lebetina*)	100%	90%	0.0
	Dis 1b	c11797_g1_i1_F	gi|48428040|sp|Q7ZZM2.1|Disintegrin jerdostatin (*Protobothrops jerdonii*)	100%	87%	3e-71
P-III 7	P-III 7a	c9453_g1_i1_M	XP_015683143.1 PREDICTED: zinc metalloproteinase-disintegrin-like NaMP (*Protobothrops mucrosquamatus*)	59%	97%	2e-13
		c66080_g1_i1_M	XP_015683143.1 PREDICTED: zinc metalloproteinase-disintegrin-like NaMP (*Protobothrops mucrosquamatus*)	100%	93%	6e-43
	P-III 7b	c22734_g1_i1_M	JAC96590.1 Snake venom metalloproteinase U (*Echis coloratus*)	100%	95%	2e-50
		c34552_g1_i1_M	JAC96590.1 Snake venom metalloproteinase U (*Echis coloratus*)	100%	92%	3e-49
		c41009_g1_i1_M	JAC96590.1 Snake venom metalloproteinase U (*Echis coloratus*)	100%	93%	2e-52
		c71069_g1_i1_M	JAC96590.1 Snake venom metalloproteinase U (*Echis coloratus*)	100%	98%	3e-54
P-II 1		c19938_g2_i1_M	sp|U5PZ28.1|VM2H1_BOTLA|Zinc metalloproteinase-disintegrin BlatH1 (*Bothriechis lateralis*)	55%	58%	2e-11
P-II 2		c11787_g3_i1_F	AAP20639.1 metalloproteinase/disintegrin saxin precursor (*Gloydius saxatilis*)	92%	78%	2e-25

Table 5.3. Summary of isoforms and expression level of SVMPs in male and female transcriptomes

Classes	Annotation	Male		Female	
		FPKM	Total FPKM per each class		Total FPKM per each class
Dis 1a	Disintegrin jerdostatin (*Protobothrops jerdonii*)	62514.52	62661.9 (75%)	18456.39	24205.66 (75%)
Dis 1b	Disintegrin jerdostatin (*Protobothrops jerdonii*)	0		5731.5	
Dis 2	VGD-containing dimeric disintegrin subunit ML-G1(*Macrovipera lebetina transmediterranea*)	82.5		10.84	
Dis 3	MLD-containing dimeric disintegrin (*Echis ocellatus*)	62.42		6.93	
Dis 4	MLD-containing dimeric disintegrin subunit ML-G2 (*Macrovipera lebetina transmediterranea*)	2.46		0	
P-III 1	factor X activator heavy chain (*Daboia russellii russellii*)	9225.14	20676.9 (25%)	3829.48	7986.59 (25%)
P-III 2	Zinc metalloproteinase-disintegrin-like VLAIP-A (*Macrovipera lebetina*)	5794.55		2115.46	
P-III 3	Snake venom metalloproteinase N (*Echis coloratus*)	5490.9		2012.55	
P-III 4	Group III snake venom metalloproteinase (Svmp3-Eoc1 gene) (*Echis ocellatus*)	139.71		25.76	
P-III 5	Cadam_SVMPIII-2e (*Crotalus adamanteus*)	5.66		3.34	
P-III 6	Cadam_SVMPIII-1a (*Crotalus adamanteus*)	14.51		0	
P-III 7a	Zinc metalloproteinase-disintegrin-like NaMP (*Protobothrops mucrosquamatus*)	2.37		0	
P-III 7b	Snake venom metalloproteinase U (*Echis coloratus*)	2.35		0	
P-II 1	Zinc metalloproteinase-disintegrin BlatH1 (*Bothriechis lateralis*)	1.41	1.41	0	8.44
P-II 2	Metalloproteinase/disintegrin saxin precursor (*Gloydius saxatilis*)	0		8.44	

5.3.2. Sequence analysis of SVMP transcripts

5.3.2.1. RVV-X

The strong abundance of P-III SVMP transcripts in the venom gland transcriptome is in agreement with a detection of RVV-X heavy chain in venomic analysis of Siamese Russell's viper (146). The Clustal alignment of translated sequences of Myanmar Russell's viper species (c20464_g1_i1_M & c11797_g1_i3_F) with those from other species showed 99% sequences identity to Indian species (ADJ67475.1); 97% identity to Indonesian species (sp|Q7LZ61.2) (97) and Thailand species (63) (N-terminal portion of the sequence is not completely available) (Figure 5.1).

```
                                 Signal peptide
D_siamensis_Indonesia       MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY   60
D_r_siamensis_Thailand      ------------------------------------------------------------   0
c20464_g1_i1_M              MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY   60
c11797_g1_i3_F              MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY   60
D_r_russellii_India         MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTAMPKGAVKQPEQKYEDTMQY   60

                                          Propeptide domain
D_siamensis_Indonesia       EFEVNGEPVVLHLEKNKILFSEDYSETHYYPDGREITTNPPVEDHCYYHGRIQNDAHSSA   120
D_r_siamensis_Thailand      --------------------SEDYSETHYYPDGREITTNPPVEDHCYYHGRIQNDAHSSA   40
c20464_g1_i1_M              EFEVNGEPVVLHLEKNKILFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIQNDAHSSA   120
c11797_g1_i3_F              EFEVNGEPVVLHLEKNKILFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIQNDAHSSA   120
D_r_russellii_India         EFEVNGEPVVLHLEKNKILFSEDYSETHYYPDGREITTNPPVEDHCYYHGHIQNDGHSSA   120
                                          ********* ********************:*****.****

                                                                  Cysteine switch
D_siamensis_Indonesia       SISACNGLKGHFKLRGEMYFIEPLKLSNSEAHAVYKYENIEKEDEIPKMCGVTQTNWESD   180
D_r_siamensis_Thailand      SISACNGLKGHFKLRGEMYFIEPLKLSNSEAHAVYKYENIEKEDEIPKMCGVTQTNWESD   100
c20464_g1_i1_M              SISACNGLKGHFKLRGEMYFIEPLKLSNNEAHAVYKYENIEKEDETPKMCGVTQTNWESD   180
c11797_g1_i3_F              SISACNGLKGHFKLRGEMYFIEPLKLSNNEAHAVYKYENIEKEDETPKMCGVTQTNWESD   180
D_r_russellii_India         SISACNGLKGHFKLRGEMYFIEPLKLSNNEAHAVYKYENIEKEDETPKMCGVTQTNWESD   180
                            ****************************.*************** ***************

D_siamensis_Indonesia       KPIKKASQLVSTSAQFNKIFIELVIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI   240
D_r_siamensis_Thailand      KPIKKASQLVSTSAQFNKIFIELVIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI   160
c20464_g1_i1_M              EPIKKASQLVATSAQFNKAFIELIIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI   240
c11797_g1_i3_F              EPIKKASQLVATSAQFNKAFIELIIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI   240
D_r_russellii_India         KPIKKASQLVSTSAQFNKAFIELIIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI   240
                            :*********:******* ****:************************************

D_siamensis_Indonesia       HVTLIGVEFWCDRDLINVTSSADETLNSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG   300
D_r_siamensis_Thailand      HVTLIGVEFWCDRDLINVTSSADETLNSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG   220
c20464_g1_i1_M              HVTLIGVEFWCDRDLINVTSSADETLDSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG   300
c11797_g1_i3_F              HVTLIGVEFWCDRDLINVTSSADETLDSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG   300
D_r_russellii_India         HVTLIGVEFWCDRDLINVTSSADETLDSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG   300
                            **************************:**********************************

                                                  Zinc binding site
D_siamensis_Indonesia       ITFLAGMCQAYRSVEIVQEQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV   360
D_r_siamensis_Thailand      ITFLAGMCQAYRSVEIVQEQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV   280
c20464_g1_i1_M              ITFLAGMCQAYRSVGIVQVQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV   360
c11797_g1_i3_F              ITFLAGMCQAYRSVGIVQVQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV   360
D_r_russellii_India         ITFLAGMCQAYRSVGIVQVQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV   360
                            ************** *** ****************************************

D_siamensis_Indonesia       LSDQPSKLFSNCSIHDYQRYLTRYKPKCIFNPPLRKDIVSPPVCGNEIWEEGEECDCGSP   420
D_r_siamensis_Thailand      LSDQPSKLFSNCSIHDYQRYLTRYKPKCIFNPPLRKDIVSPPVCGNEIWEEGEECDCGSP   340
c20464_g1_i1_M              LSDQPSKLFSNCSIHDYQRYLTRYKPKCILYPPLRKDIVSPPVCGNEIWEEGEECDCGSP   420
c11797_g1_i3_F              LSDQPSKLFSNCSIHDYQRYLTRYKPKCILYPPLRKDIVSPPVCGNEIWEEGEECDCGSP   420
D_r_russellii_India         LSDQPSKLFSNCSIHDYQRYLTRYKPKCILYPPLRKDIVSPPVCGNEIWEEGEECDCGSP   420
                            ****************************: ****************************

                                     Disintergrin-like domain
D_siamensis_Indonesia       ANCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARDECDVPEHCTGQSAECP   480
D_r_siamensis_Thailand      ANCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARDECDVPEHCTGQSAECP   400
c20464_g1_i1_M              ADCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARNECDVPEHCTGQSAECP   480
c11797_g1_i3_F              ADCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARNECDVPEHCTGQSAECP   480
D_r_russellii_India         ADCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARNECDVPEHCTGQSAECP   480
                            *:*****************************************:****************

D_siamensis_Indonesia       RDQLQQNGKPCQNNRGYCYNGDCPIMRNQCISLFGSRANVAKDSCFQENLKGSYYGYCRK   540
D_r_siamensis_Thailand      RDQLQQNGKPCQNNRGYCYNGDCPIMRNQCISLFGSRANVAKDSCFQENLKGSYYGYCRK   460
c20464_g1_i1_M              RDQLQQNGQPCQNNRGYCYNGDCPIMRNQCISLFGSRANVAKDSCFQENLKGSYYGYCRK   540
c11797_g1_i3_F              RDQLQQNGQPCQNNRGYCYNGDCPIMRNQCISLFGSRANVAKDSCFQENLKGSYYGYCRK   540
D_r_russellii_India         RDQLQQNGQPCQNNRGYCYNGDCPIMRNQCISLFGSRATVAKDSCFQENLKGSYYGYCRK   540
                            ********:****************************.*********************
```

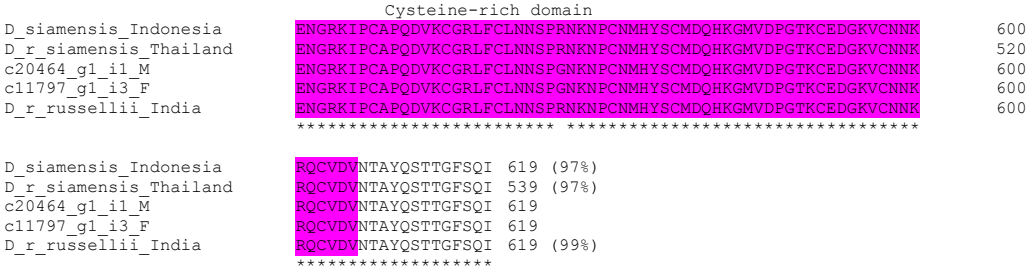


Figure 5.1. Clustal alignment of RVV-X heavy chains from different viper species. Signal peptide, Cysteine switch motif and zinc binding site are underlined. Prepropeptide region is in blue green colour. Different domains are highlighted in different colours: Metalloproteinase domain in yellow, disintegrin-like domain in green, cysteine-rich domain in pink. Letters corresponding to the varied amino acid residues are shown in reds.

Since the heavy chain of RVV-X functions cooperatively with light chains, RVV-X light chain transcripts (c20336_g1_i7_M & c20336_g1_i9_M) from MRV transcriptome have also been retrieved and aligned with the light chains of different species.

RVV-X light chains are C-type lectin-like proteins which are linked to the heavy chain through disulfide bonds (95). The two light chains recognise the Ca^{2+}-bound conformation of the Gla domain in factor X and help in the protease action of the heavy chain on factor X (29). The alignment of light chain1 showed that light chain 1 transcript of MRV has a 97% similarity to that of Thailand (63), India (ADJ67474.1), Myanmar (ADK22820.1) species and Eastern Russell's viper [Q4PRD1 (SLLC1_DABSI)]; and 98% identity to Indonesian (97) species.

It is noticeable that the glutamate residues next to the Ca^{2}+ binding site and in the C-terminal region in protein sequences from Thailand, Indonesia and Eastern Russell's viper species are substituted by lysine and valine, consequently, in Myanmar and India species (Figure 5.2). This difference in side chain polarity might be involved in the functional property of the proteins. The primary capture site for Factor X in the RVV-X is formed by the concave cleft between the two light chains. The interaction of the Gla domain-binding site (formed by two light chains) of RVV-X with the Gla domain of Factor X, plays an essential role in Ca^{2+}-dependent activation of Factor X by RVV-X. Thus, the physiochemical properties of amino acids which are involved the

formation of both the Gla domain binding site and the Ca^{2}+-binding sites might play an important role in RVV-X protein folding structure, as well as the interaction between Factor X (substrate) and RVV-X (enzyme). In the future, detailed interactions of those active residues should be further studied by constructing a model with the QUANTA program.

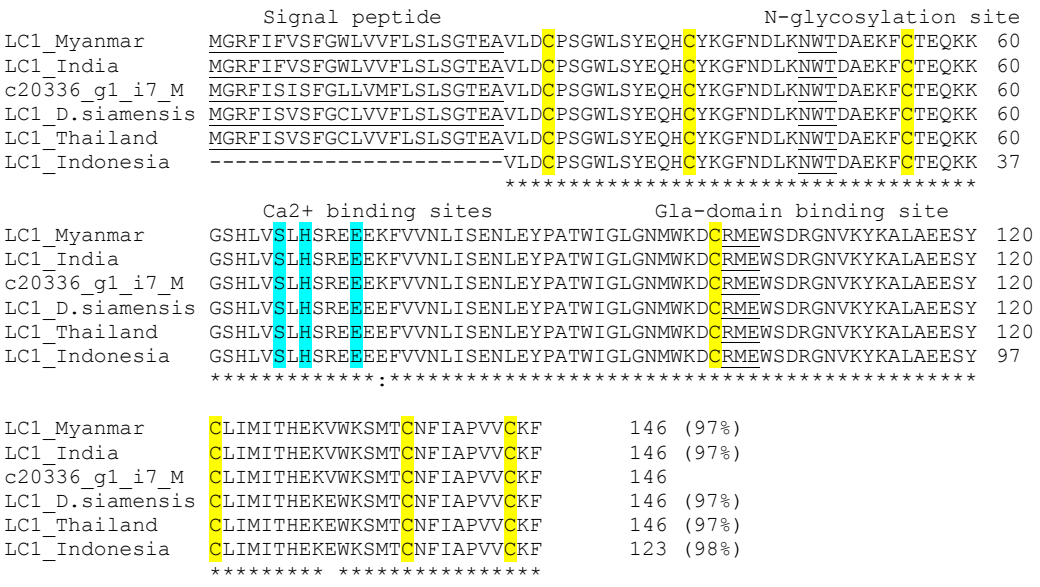


Figure 5.2. Sequence alignment of RVV-X-Light chain 1 proteins from different viper species. Glutamic acid (E) (red colour) is present in proteins from Thailand species, Lysine (K) (green colour) in proteins from Indonesian species and Eastern Russell's viper are changed to and Valine (V) (Purple colour) in Myanmar and Indian species.

The translated amino acid sequences of the light chain 2 transcript of MRV has 100% identity to those of Eastern Russel's viper [Q4PRD2 (SLLC2_DABSI)] and Myanmar (ADK22819.1) species, 96% to Taiwan (AFE61611.1) species, 93% to Indonesian (97) species, and 89% to that of Indian (ADJ67473.1) species transcripts. A protein alignment of light chain 2 shows variation in amino acid residues among the different species. This reflects the particular functional changes in light chain 2 proteins from different regional viper species (Figure 5.3). Amino acid changes not reflected in the physiochemical properties of a protein are shown in red, yellow and blue colours. Amino acids with different physiochemical properties are highlighted in magenta and

green colours. Thus, these latter sequence variations among Russell's viper from different regions may explain the diverse clinical manifestations of snake bite patients in different countries.

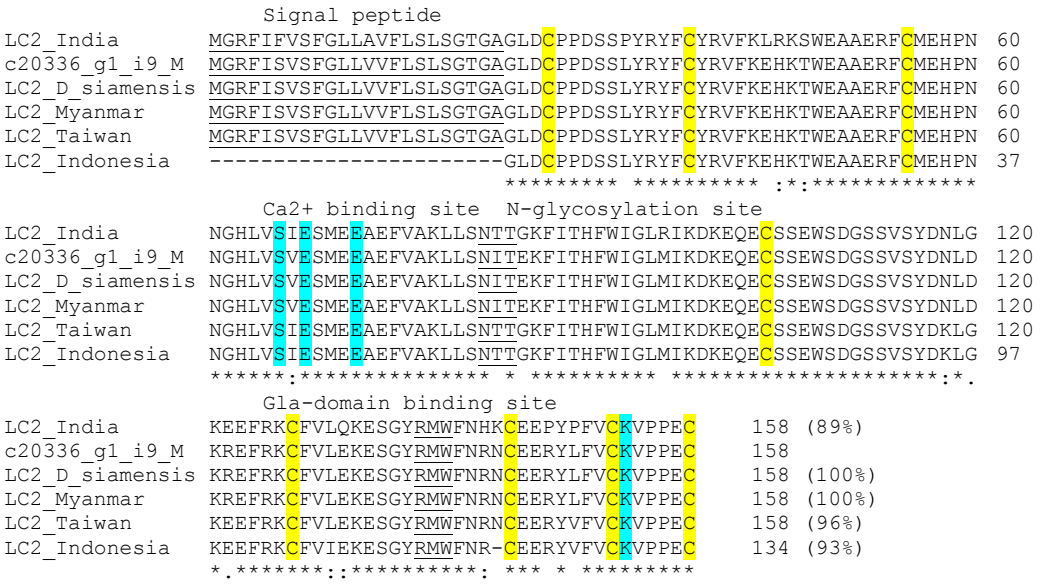


Figure 5.3. Sequence alignment of RVV-X-Light chain 2 proteins from different viper species. The amino acid residues having same physiochemical properties are shown in red, yellow and blue colours. The amino acids having different physiochemical properties are in magenta and green colours.

5.3.2.2. Disintegrins

Snake venom disintegrins are generated by proteolytic processing of P-II SVMP precursors (86) or are synthesized from short coding mRNA without the metalloproteinase-coding region (147). In the current venom gland transcriptome, disintegrin transcripts are the most abundantly expressed among SVMP transcripts in both male and female sample. One disintegrin isoform (Dis 1a) was highly expressed in both male and female venom glands and another isoform (Dis 1b) was selectively expressed in the female-one. The Dis 1a and Dis 1b proteins were found to possess amino acid variations in the signal peptide and N-terminus of the coding region with functional disintegrin motifs being conserved (Figure 5.4). The Dis 1a sequence is identical to the sequence of clone 1.2.2.14 from the cDNA library

mentioned in chapter 4. The Dis 1a sequence was found to be identical the russellistatin from *D. russelii* as well as to jerdostatin from *Trimeresurus jerdonii*. The disintegrin domain of all 3 sequences contain RTS motif. Sanz-Soler R et al (2012) showed that evolutionary histories of the RTS/KTS short disintegrins are independent from a phylogeny of RGD short integrins (148) .

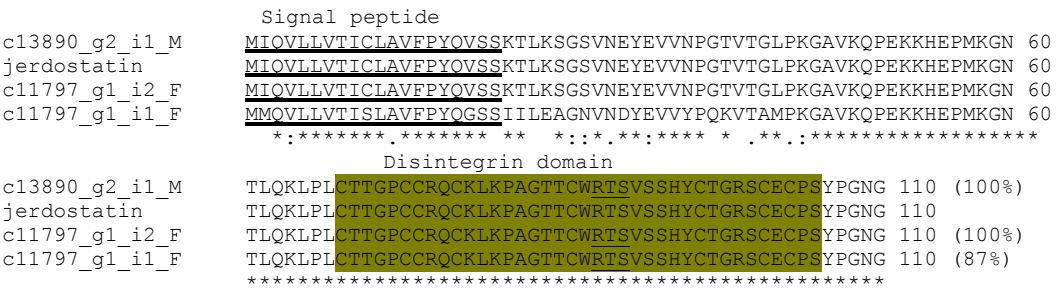


Figure 5.4. Clustal alignment of translated disintegrin transcripts from MRV with that of jerdostatin (disintegrin from *Trimeresurus jerdonii*) with conserved integrin binding motif RTS. Isoform 1 (c13890_g2_i1_M and c11797_g1_i2_F) protein sequence has 100% identity and isoform 2 (c11797_g1_i1_F) sequence has 87% identity to jerdostatin sequence.

5.3.2.3. *V. lebetina* apoptosis-inducing protease-A (VLAIP-A) homolog

The second most highly expressed P-III SVMP transcripts, c20464_g1_i2_M & c11797_g1_i4_F, in both male and female transcriptomes are homologous to Zinc metalloproteinase-disintegrin-like VLAIP-A from *Macrovipera lebetina* (Figure 5.5). These transcripts are partially sequenced but contain regions encoding signal peptide, prepro-peptide, and metalloproteinase, disintegrin-like and cysteine-rich protein domains. Our deduced amino acid sequences have the critical functional motifs that belonged to VLAIP-A as follows: the cysteine switch motif (PKMCGV) is in the prodomain and the Zn^{2+}-chelating sequence HEXXHXXGXXHD in the metalloproteinase domain. The exact sequence of Zn-binding site **HE**MG**H**NL**G**ME**HD** is different from that of RVV-X **HE**LS**H**NL**G**MY**HD** due to some amino acid variation. The sequence also possesses the conserved intrachain disulfide bond pattern of 6-

Cys: Cys314-Cys394, Cys354-Cys378 and Cys356-Cys361 were found. The typical Cys369 residue for vascular apoptosis inducing proteases was also involved. The disintegrin-like domain was found to contain the ECD motif for binding to integrin receptor. The VLAIP from *M. lebetina* showed proteolytic activity on azocasein, fibrinogen (readily to the Aα-chain and more slowly to Bβ-chain) and oxidized insulin B-chain. The VLAIP also inhibits endothelial cell adhesion to extracellular matrix proteins: fibrinogen, fibronectin, vitronectin, collagen I and collagen IV. The primary effects of VLAIP on HUVECs (human umbilical vein endothelial cells) were induction of changes in the cell shape and in the attachment of cells to the substrate followed by their detachment, subsequent aggregation and cell death (149). Since the VLAIP homolog is a very new finding in regards to the Myanmar Russell's viper species, it will be named as *Daboia siamensis* apoptosis-inducing protease (DSAIP) from hereon.

```
                    Signal peptide
c20464_g1_i2_M      MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY 60
c11797_g1_i4_F      MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY 60
VLAIP-A             MMQVLLVTISLAVFPYQGSSIILESGNVNDYEVVYPQKVTAMPKGAVKQPEQKYEDAMQY 60
                    ****************************************:*****:********:***
                    Prepropeptide
c20464_g1_i2_M      EFEVNGEPVVLHLEKNKILFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIQNDAHSSA 120
c11797_g1_i4_F      EFEVNGEPVVLHLEKNKILFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIQNDAHSSA 120
VLAIP-A             EFKVKGEPVVLLLEKNKDLFSEDYSETHYSPDGREITTNPPVEDHCYYHGRIQNDADSSA 120
                    **:*:****** ***** ************************************.***
                                                    Cysteine switch
c20464_g1_i2_M      SISACNGLKGHFKLRGEMYFIEPLKLSNNEAHAVYKYENIEKEDETPKMCGVTQTNWESD 180
c11797_g1_i4_F      SISACNGLKGHFKLRGEMYFIEPLKLSNNEAHAVYKYENIEKEDETPKMCGVTQTNWESD 180
VLAIP-A             SISACNGLKGHFMLQGETYLIEPLKLPDSEAHAVYKYENVEKEDEAPKMCGVTQTNWESD 180
                    ************ *:** *:******.:.**********:*****:**************

c20464_g1_i2_M      EPIKKASLLNLTPEQRTYLKSKKYIEVIIIADYIIYLKYGRNLFTIRTRIYEIVNILNVI 240
c11797_g1_i4_F      EPIKKASLLNLTPEQRTYLKSKKYIEVIIIADYIIYLKYGRNLFTIRTRIYEIVNILNVI 240
VLAIP-A             EPIKKASQLNLTPEQRRYLNSPKYIKLVIVADYIMFLKYGRSLITIRTRIYEIVNILNVI 240
                    ******* ******** **:* ***:::*:****::*****.*:****************

c20464_g1_i2_M      YRVLNIYIALLGLEIWNNGDKINVLPEAKVTLDLFGKWKASDLLNHRNHDNAYLLTGINF 300
c11797_g1_i4_F      YRVLNIYIALLGLEIWNNGDKINVLPEAKVTLDLFGKWKASDLLNHRNHDNAYLLTGINF 300
VLAIP-A             YRVLNIYIALLGLEIWNNGDKINVLPETKVTLDLFGKWRERDLLNRRKHDNAQLLTDINF 300
                    ***************************:**********:  ****:*:**** ***.***
                                                  Zn-binding site
c20464_g1_i2_M      NGPTAGLGYLGGMCHPEYSAGIVQDHNKLNYLVALAMAHEMGHNLGMEHDKIHCTCGAKS 360
c11797_g1_i4_F      NGPTAGLGYLGGMCHPEYSAGIVQDHNKLNYLVALAMAHEMGHNLGMEHDKIHCTCGAKS 360
VLAIP-A             NGPTAGLGYVGSMCDPQYSAGIVQDHNKVNFLVALAMAHEMGHNLGMEHDEIHCTCGAKS 360
                    **********:*.**.*:***********:*:********************:*********

c20464_g1_i2_M      CIMSGTLSCEASNRFSDCSREEHQKYLIDKMPQCILNKPLKTDIVSPAVCGNYLVELGED 420
c11797_g1_i4_F      CIMSGTLSCEASNRFSDCSREEHQKYLIDKMPQCILNKPLKTDIVSPAVCGNYLVELGED 420
VLAIP-A             CIMSGTLSCEASIRFSNCSREEHQKYLINKMPQCILNKPLKTDIVSPAVCGNYLVELGED 420
                    ************ ***:***********:******************************
                                                        Disintegrin-like domain
c20464_g1_i2_M      CDCGSPSVCQNPCCNAATCKLTQGSQCADEECCDQCKFRRAGTVCRAANGECDVSDLCTG 480
c11797_g1_i4_F      CDCGSPSVCQNPCCNAATCKLTQGSQCADEECCDQCKFRRAGTVCRAANGECDVSDLCTG 480
VLAIP-A             CDCGSPRDCQNPCCNAATCKLTPGSQCADGECCDQCKFRRAGTVCRPANGECDVSDLCTG 480
                    ******  ************** ****** ****************.*************
```

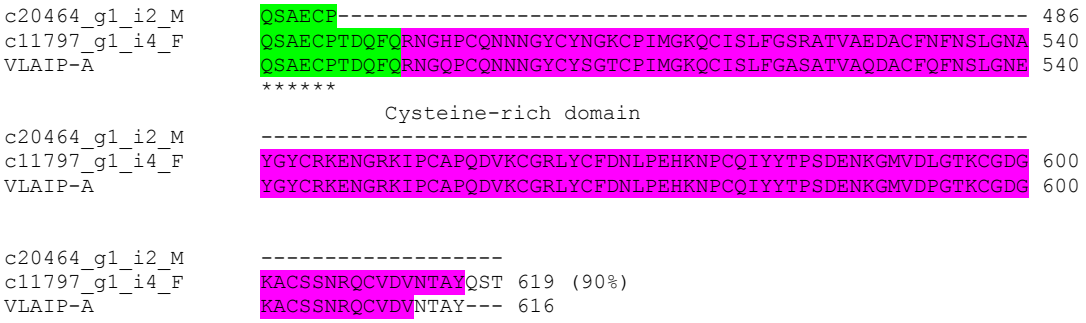


Figure 5.5. Clustal alignment of translated transcripts (partial-length CDS) of VLAIP-A homolog from MRV with that of zinc metalloproteinase-disintegrin-like VLAIP-A (*V. lebetina* apoptosis-inducing protease from *Macrovipera lebetina*). The shorter translated sequence from male sample, c20464_g1_i2_M has 486 amino acid residues. The longer translated sequence from female sample, c11797_g1_i4_F (619 amino acid residues) has 90% identity to VLAIP-A sequence. Signal peptide, Cysteine switch motif, and zinc binding site are underlined. The prepropeptide region is in blue green. Different domains are highlighted in different colours: Metalloproteinase domain is shown in yellow, disintgerin-like domain in green, and cysteine-rich domain in pink. The Cysteine residues are in red letters.

5.3.2.4. Retrotransposable sequences (SINEs)

SINEs (short interspersed elements) are non-viral retrotransposable repetitive sequences and are relatively abundant in advanced snakes' genome (150). Retrotransposition of SINEs depend on reverse transcriptase and endonuclease activities encoded by LINEs (long interspersed elements). The Sauria SINEs in snake genome were identified by Piskurek et al., (2006) and they pointed out that these retrotransponsons can act like genetic markers for reptile evolution (151). The venomous copperhead (*Agkistrodon contortrix*) was found to have 23-fold greater levels of transposable element (TE)-related transcripts than the Burmese python (*Python molurus bivittatus*) suggesting that the TE might facilitate the expansion and regulatory process of venom gene families in venomous snakes (152). The presence of long interspersed sequences (LINEs) inside the PLA_2 genes and in Viperidae snake genomes suggested that the involvement of retrotransposable elements on PLA_2

gene duplication (153). In *Prothrobothrops flavoviridis* (habu snake), an LINE1 (L1), named PfL1 retrotransposable element was found in the intronic sequence of a fetuin-encoding gene. That genomic segment retains an open reading frame (ORF) that encodes a reverse transcriptase (RT)-like protein (PfRT).

Moreover, the habu snake liver RNA indicated active transcription of the PfRT. This suggests that sufficient RT activity could contribute to the accelerated evolution of exonic nucleotide sequences in the genes for venom proteins (154). Likewise, SINE in SVMPs transcripts may play a role in generation of SVMPs gene diversity. SINEs can act as regulators of gene expression as key elements together with LINEs in the recruitment and amplification process of SVMPs into the venom gland (69). However, the molecular mechanism for transposon involvement in evolution of toxin gene families is still elusive.

5.3.3. Molecular cloning of SVMPs

5.3.3.1. RVV-X heavy chain

5.3.3.1.1. Primer design and cloning

The primers were designed from sequences of contig c20464_g1_i1_M (1900 bp) and c11797_g1_i3_F (2300 bp). The forward primer is in the region just before the start codon and the reverse primer is in the region after stop codon.

Primer	Sequence	nt	Tm (°C)	GC %
Forward primer	5'- AAATCCGCCCTCCAAAATGATGC -3'	23	62.9	53%
Reverse primer	5'-TGAACCTTCTGGAAGAAAGATCTCC -3'	25	64.2	44%

The first strand cDNA was synthesised from total RNA of the venom gland of female using Oligo (dT)18 primer by using RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific). The PCR amplification of 1st strand cDNA was performed using gene specific forward and reverse primers using FastStart Taq DNA Polymerase (Roche). The condition for PCR was as follows: initial denaturation for 2 mins at 94°C, followed by 35 extension cycles (denaturation for 40 sec at 94°C, annealing for 30 sec

at 52°C, and elongation for 2 min at 72°C), and a terminal extension step of 72°C for 5 min.

Figure 5.6. Amplified PCR products of RVV-X heavy chain sequence. It showed 2 bands: 1900 bp and 500 bp on gel electrophoresis.

The 2 bands (Figure 5.6) were cut and purified using Wizard SV Gel and PCR Clean-Up System (Promega). The purified PCR product of 500 bp band was then amplified with Big Dye Terminator v3.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems).

The sequencing result showed that the sequence was 99% identity to the jerdostatin mRNA from *Trimeresurus jerdonii* (e-value = $2x10^{-156}$).

To obtain a PCR product of 1900 bp alone, PCR was performed again with gradient temperature (50-60 °C). The 1900-bp PCR products were isolated at annealing temperatures of 58°C and 60°C (Figure 5.7).

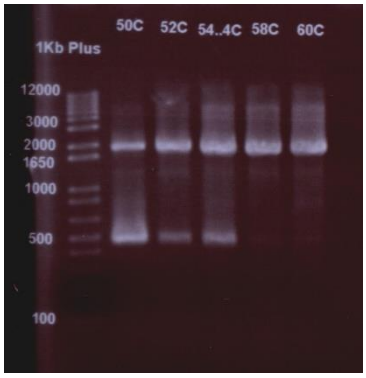


Figure 5.7. Amplification products of RVV-X heavy chain gene using gradient temperature condition (50°C to 60°C).

PCR products at 58°C were purified using Wizard®SV Gel and PCR Clean-Up System before ligated into pGEM-T Easy vector. The ligated plasmids were then transformed into JM109 competent cells in the transformation tube by a heat-shock method. The transformed cells were culture on LB/Ampicillin/IPTG/X-Gal agar plates. The white colonies were selected and cloned in LB medium in 37°C shaking incubator at 225 rpm overnight. The plasmids were then purified using GF-1 Plasmid DNA Extraction kit. The cDNA insert was identified on agarose gel (1%) after digested with the EcoRI restriction enzyme (Figure 5.8).

Figure 5.8. Identification of cDNA insert in pGEM-T Easy vector after incubation of plasmids with EcoRI at 37°C for 4 hours. The cut plasmid sample showed 2000 bp cDNA insert.

5.3.3.1.2. DNA sequencing

Firstly, DNA sequencing with M13pUC Forward and M13pUC Reverse primers was performed. Since the length of gene is more than 800 bp, the walking primers were then designed to get the whole sequence.

Primer	Sequence	nt	Tm	GC%
11797R1	5’-ACT GCT CCT TTG GGC AAT GC-3’	20	60.5˚C	55%
11797F1	5’-TCA TCT GCA AGC ATC AGT GC-3’	20	58.4 ˚C	50%
11797F2	5’-TGC GAC AGA GAT TTG ATT AAC G-3’	22	58.4 ˚C	41%
11797F3	5’-CAA AAT GCA TTC TCT ATC CAC C-3’	22	58.4 ˚C	41%

```
CLUSTAL O(1.2.4) multiple sequence alignment

c11797_g1_i3_F      CATTCCTTGCTTCTCATTGTCAAATGAGGAAGAGCTCCTGTTGGCTTGAAAGCAGGAAGA  60
Seq                 ------------------------------------------------------------G  1

                     11797R1
c11797_g1_i3_F      GATTGCCAGCCAAATCCGGCCTCCAAAATGATGCAAGTTCTCTTAGTAACTATAAGCTTA  120
Seq                 GGAATTCGATTAAATCCGCCCTCCAAAATGATGCAAGTTCTCTTAGTAACTATAAGCTTA  61
                    *     *    ******* ************************************

                                                          11797_M13F
c11797_g1_i3_F      GCAGTTTTTCCATATCAAGGGAGCTCTATAATCCTGGAATCTGGGAACGTTAATGATTAT  180
Seq                 GCAGTTTTTCCATATCAAGGGAGCTCTATAATCCTGGAATCTGGGAACGTTAATGATTAT  121
                    ************************************************************

c11797_g1_i3_F      GAAGTAGTGTATCCACAAAAAGTCACTGCATTGCCCAAAGGAGCAGTTCAGCAGCCTGAG  240
Seq                 GAAGTAGTGTATCCACAAAAAGTCACTGCATTGCCCAAAGGAGCAGTTCAGCAGCCTGAG  181
                    ************************************************************

c11797_g1_i3_F      CAAAAGTATGAAGACACCATGCAATATGAATTTGAAGTGAATGGAGAGCCAGTGGTCCTT  300
Seq                 CAAAAGTATGAAGACACCATGCAATATGAATTTGAAGTGAATGGAGAGCCAGTGGTCCTT  241
                    ************************************************************

c11797_g1_i3_F      CACCTGGAAAAAAATAAAATACTTTTTTCAGAAGATTACAGCGAGACTCATTATTCCCCA  360
Seq                 CACCTGGAAAAAAATAAAATACTTTTTTCAGAAGATTACAGCGAGACTCATTATTCCCCA  301
                    ************************************************************

c11797_g1_i3_F      GATGGCAGAGAAATTACAACAAACCCTCCGGTTGAGGATCACTGCTATTATCATGGACGC  420
Seq                 GATGGCAGAGAAATTACAACAAACCCTCCGGTTGAGGATCACTGCTATTATCATGGACGC  361
                    ************************************************************

c11797_g1_i3_F      ATCCAGAATGATGCTCACTCATCTGCAAGCATCAGTGCATGCAATGGTTTGAAAGGACAT  480
Seq                 ATCCAGAATGATGCTCACTCATCTGCAAGCATCAGTGCATGCAATGGTTTGAAAGGACAT  421
                    ************************************************************

                                                              11797F1
c11797_g1_i3_F      TTCAAGCTTCGAGGGGAGATGTACTTTATTGAACCCTTGAAGCTTTCCAACAATGAAGCC  540
Seq                 TTCAAGCTTCGAGGGGAGATGTACTTTATTGAACCCTTGAAGCTTTCCAACAATGAAGCC  481
                    ************************************************************

c11797_g1_i3_F      CATGCAGTCTACAAATATGAAAACATAGAAAAAGAGGATGAGACCCCCAAAATGTGTGGG  600
Seq                 CATGCAGTCTACAAATATGAAAACATAGAAAAAGAGGATGAGACCCCCAAAATGTGTGGG  541
                    ************************************************************

c11797_g1_i3_F      GTAACCCAGACTAATTGGGAATCAGATGAGCCCATCAAAAAGGCCTCTCAGTTAGTTGCT  660
Seq                 GTAACCCAGACTAATTGGGAATCAGATGAGCCCATCAAAAAGGCCTCTCAGTTAGTTGCT  601
                    ************************************************************

c11797_g1_i3_F      ACTTCTGCACAATTCAACAAAGCATTCATTGAGCTTATCATAATTGTGGACCACAGCATG  720
Seq                 ACTTCTGCACAATTCAACAAAGCATTCATTGAGCTTATCATAATTGTGGACCACAGCATG  661
                    ************************************************************

c11797_g1_i3_F      GCCAAGAAATGCAATTCAACTGCTACAAATACAAAAATATATGAAATTGTCAACAGTGCA  780
```

```
Seq              GCCAAGAAATGCAATTCAACTGCTACAAATACAAAAATATATGAAATTGTCAACAGTGCA  721
                 ************************************************************

c11797_g1_i3_F   AATGAGATTTTTAATCCTTTGAATATTCATGTAACATTGATTGGTGTAGAATTTTGGTGC  840
Seq              AATGAGATTTTTAATCCTTTGAATATTCATGTAACATTGATTGGTGTAGAATTTTGGTGC  781
                 ************************************************************

c11797_g1_i3_F   GACAGAGATTTGATTAACGTGACATCATCAGCAGATGAAACTTTGGACTCATTTGGAGAA  900
Seq              GACAGAGATTTGATTAACGTGACATCATCAGCAGATGAAACTTTGGACTCATTTGGAGAA  841
                 ************************************************************

c11797_g1_i3_F   TGGAGAGCGTCAGATTTGATGACTCGGAAAAGCCATGATAATGCTCTGTTATTCACGGAC  960
Seq              TGGAGAGCGTCAGATTTGATGACTCGGAAAAGCCATGATAATGCTCTGTTATTCACGGAC  901
                 ************************************************************

c11797_g1_i3_F   ATGAGATTCGATTTAAACACTTTGGGAATCACTTTCTTAGCTGGCATGTGCCAGGCATAT  1020
Seq              ATGAGATTCGATTTAAACACTTTGGGAATCACTTTCTTAGCTGGCATGTGCCAGGCATAT  961
                 ************************************************************

                   11797F2
c11797_g1_i3_F   CGTTCTGTAGGAATTGTTCAGGTACAAGGGAACAGAAATTTTAAGACTGCAGTTATAATG  1080
Seq              CGTTCTGTAGGAATTGTTCAGGTACAAGGGAACAGAAATTTTAAGACTGCAGTTATAATG  1021
                 ************************************************************

c11797_g1_i3_F   GCCCATGAGCTGAGTCATAATCTGGGCATGTATCATGACGGAAAAAACTGTATTTGTAAT  1140
Seq              GCCCATGAGCTGAGTCATAATCTGGGCATGTATCATGACGGAAAAAACTGTATTTGTAAT  1081
                 ************************************************************

c11797_g1_i3_F   GATTCCTCATGTGTTATGTCTCCTGTGCTAAGCGATCAACCTTCCAAATTGTTCAGCAAT  1200
Seq              GATTCCTCATGTGTTATGTCTCCTGTGCTAAGCGATCAACCTTCCAAATTGTTCAGCAAT  1141
                 ************************************************************

c11797_g1_i3_F   TGTAGTATTCACGATTATCAGAGGTATCTTACTAGATATAAACCAAAATGCATTCTCTAT  1260
Seq              TGTAGTATTCACGATTATCAGAGGTATCTTACTAGATATAAACCAAAATGCATTCTCTAT  1201
                 ************************************************************

c11797_g1_i3_F   CCACCCTTGAGAAAAGATATTGTTTCACCTCCCGTTTGCGGAAATGAAATTTGGGAGGAG  1320
Seq              CCACCCTTGAGAAAAGATATTGTTTCACCTCCCGTTTGCGGAAATGAAATTTGGGAGGAG  1261
                 ************************************************************

                  Disintegrin-like domain
c11797_g1_i3_F   GGAGAAGAATGTGACTGTGGCTCTCCTGCAGATTGTCAAAATCCGTGCTGTGATGCTGCA  1380
Seq              GGAGAAGAATGTGACTGAGGCTCTCCTGCAGATTGTCAAAATCCGTGCTGTGATGCTGCA  1321
                 ********** ***************************************************
                            Stop codon

                       11797F3
c11797_g1_i3_F   ACATGTAAACTGAAGCCAGGGGCAGAGTGTGGAAATGGACTGTGTTGTTACCAATGCAAA  1440
Seq              ACATGTAAACTGAAGCCAGGGGCAGAGTGTGGAAATGGACTGTGTTGTTACCAATGCAAA  1381
                 ************************************************************

c11797_g1_i3_F   ATTAAGACAGCAGGAACAGTATGCCGGAGAGCAAGGAATGAGTGTGACGTCCCTGAACAC  1500
Seq              ATTAAGACAGCAGGAACAGTATGCCGGAGAGCAAGGAATGAGTGTGACGTCCCTGAACAC  1441
                 ************************************************************

c11797_g1_i3_F   TGCACTGGCCAATCTGCTGAGTGTCCCAGAGATCAGTTGCAACAGAATGGACAACCATGC  1560
Seq              TGCACTGGCCAATCTGCTGAGTGTCCCAGAGATCAGTTGCAACAGAATGGACAACCATGC  1501
                 ************************************************************

c11797_g1_i3_F   CAAAACAACAGAGGTTATTGCTACAATGGGGATTGCCCCATCATGAGAAACCAATGTATT  1620
Seq              CAAAACAACAGAGGTTATTGCTACAATGGGGATTGCCCCATCATGAGAAACCAATGTATT  1561
                 ************************************************************

c11797_g1_i3_F   TCTCTCTTTGGGTCACGTGCAAATGTGGCTAAAGATTCATGTTTTCAGGAAAACCTGAAG  1680
Seq              TCTCTCTTTGGGTCACGTGCAAATGTGGCTAAAGATTCATGTTTTCAGGAAAACCTGAAG  1621
                 ************************************************************

c11797_g1_i3_F   GGCAGTTATTATGGCTACTGCAGAAAGGAAAATGGTAGAAAGATTCCATGTGCACCACAA  1740
Seq              GGCAGTTATTATGGCTACTGCAGAAAGGAAAATGGTAGAAAGATTCCATGTGCACCACAA  1681
                 ************************************************************

c11797_g1_i3_F   GATGTAAAATGTGGCAGGTTATTCTGCTTAAATAATTCACCTGGAAACAAGAATCCTTGC  1800
Seq              GATGTAAAATGTGGCAGGTTATTCTGCTTAAATAATTCACCTGGAAACAAGAATCCTTGC  1741
                 ************************************************************

                                                        11797_M13R
c11797_g1_i3_F   AACATGCACTATAGCTGCATGGATCAACATAAGGGAATGGTTGACCCTGGAACAAAATGT  1860
Seq              AACATGCACTATAGCTGCATGGATCAACATAAGGGAATGGTTGACCCTGGAACAAAATGT  1801
                 ************************************************************

c11797_g1_i3_F   GAAGATGGAAAGGTCTGCAACAACAAAAGGCAGTGTGTTGATGTGAATACAGCCTACCAA  1920
Seq              GAAGATGGAAAGGTCTGCAACAACAAAAGGCAGTGTGTTGATGTGAATACAGCCTACCAA  1861
```

Figure 5.9. Alignment of RVV-X heavy chain sequence from NGS and that from Sanger sequencing using walking primers. The start and stop codon of sequence from NGS sequencing are in red. The start codon of sequence from Sanger sequencing is in red and the stop codon is in green. The point mutated nucleotide is in yellow.

The alignment of RVV-X heavy chain sequence from Next-generation sequencing and that obtained from Sanger sequencing with walking primers is shown in Figure 5.9. All the nucleotide sequences were identical except for the nucleotide 'T' at 1238 position of contig (yellow colour) which was expressed as 'A' in Sanger sequencing result. This variation leads to a nonsense mutation as codon 'TGT' for cysteine changed to 'TGA' for stop codon. This mutation leads to the formation of a truncated protein. This kind of protein truncation caused by the existence of a stop codon in the disintegrin-like domain region of a P-III SVMP was also found in a corresponding sequence from *Phalotris mertensi* transcriptome (155). In the later cDNA sequence, shortly after the stop codon, an-unknown sequence replaced the remaining 3'UTR region, indicating that exon 11 of the P-III SVMP gene was substituted with a new sequence. In contrast, in our cDNA, the remaining part of the transcript shortly after the stop codon is the same as the sequence of RVV-X heavy chain (Figure 5.10).

The evolution of SVMP classes is supposed to have happened through changes in surface-exposed residues of toxins and a loss of introns in several independent events within the Viperidae family (68, 156). This observation of a shortening of an original protein sequence by insertion of a stop codon is the very first in regards to Russell's viper species venom transcriptome. Thus, the acquisition of P-I and P-II types of SVMPs from P-III ones in Russell's viper might have occurred through point mutation as well as through a loss of introns.

```
DNA:      gggaattcgattaaatccgccctccaaaatgatgcaagttctcttagtaactataagctt 60
Protein:   ·G··I··R··L··N··P··P··S··K··M··M··Q··V··L··L··V··T··I··S··L·

DNA:      agcagtttttccatatcaagggagctctataatcctggaatctgggaacgttaatgatta 120
Protein:   ·A··V··F··P··Y··Q··G··S··S··I··I··L··E··S··G··N··V··N··D··Y·

DNA:      tgaagtagtgtatccacaaaaagtcactgcattgcccaaaggagcagttcagcagcctga 180
Protein:   ·E··V··V··Y··P··Q··K··V··T··A··L··P··K··G··A··V··Q··Q··P··E·

DNA:      gcaaaagtatgaagacaccatgcaatatgaatttgaagtgaatggagagccagtggtcct 240
Protein:   ·Q··K··Y··E··D··T··M··Q··Y··E··F··E··V··N··G··E··P··V··V··L·

DNA:      tcacctggaaaaaaataaaatacttttttcagaagattacagcgagactcattattcccc 300
Protein:   ·H··L··E··K··N··K··I··L··F··S··E··D··Y··S··E··T··H··Y··S··P·

DNA:      agatggcagagaaattacaacaaaccctccggttgaggatcactgctattatcatggacg 360
Protein:   ·D··G··R··E··I··T··T··N··P··P··V··E··D··H··C··Y··Y··H··G··R·

DNA:      catccagaatgatgctcactcatctgcaagcatcagtgcatgcaatggtttgaaaggaca 420
Protein:   ·I··Q··N··D··A··H··S··S··A··S··I··S··A··C··N··G··L··K··G··H·

DNA:      tttcaagcttcgaggggagatgtactttattgaacccttgaagctttccaacaatgaagc 480
Protein:   ·F··K··L··R··G··E··M··Y··F··I··E··P··L··K··L··S··N··N··E··A·

DNA:      ccatgcagtctacaaatatgaaaacatagaaaaagaggatgagacccccaaaatgtgtgg 540
Protein:   ·H··A··V··Y··K··Y··E··N··I··E··K··E··D··E··T··P··K··M··C··G·

DNA:      ggtaacccagactaattgggaatcagatgagcccatcaaaaaggcctctcagttagttgc 600
Protein:   ·V··T··Q··T··N··W··E··S··D··E··P··I··K··K··A··S··Q··L··V··A·

DNA:      tacttctgcacaattcaacaaagcattcattgagcttatcataattgtggaccacagcat 660
Protein:   ·T··S··A··Q··F··N··K··A··F··I··E··L··I··I··I··V··D··H··S··M·

DNA:      ggccaagaaatgcaattcaactgctacaaatacaaaaatatatgaaattgtcaacagtgc 720
Protein:   ·A··K··K··C··N··S··T··A··T··N··T··K··I··Y··E··I··V··N··S··A·

DNA:      aaatgagatttttaatcctttgaatattcatgtaacattgattggtgtagaatttggtg 780
Protein:   ·N··E··I··F··N··P··L··N··I··H··V··T··L··I··G··V··E··F··W··C·

DNA:      cgacagagatttgattaacgtgacatcatcagcagatgaaactttggactcatttggaga 840
Protein:   ·D··R··D··L··I··N··V··T··S··S··A··D··E··T··L··D··S··F··G··E·

DNA:      atggagagcgtcagatttgatgactcggaaaagccatgataatgctctgttattcacgga 900
Protein:   ·W··R··A··S··D··L··M··T··R··K··S··H··D··N··A··L··L··F··T··D·

DNA:      catgagattcgatttaaacactttgggaatcactttcttagctggcatgtgccaggcata 960
Protein:   ·M··R··F··D··L··N··T··L··G··I··T··F··L··A··G··M··C··Q··A··Y·
```

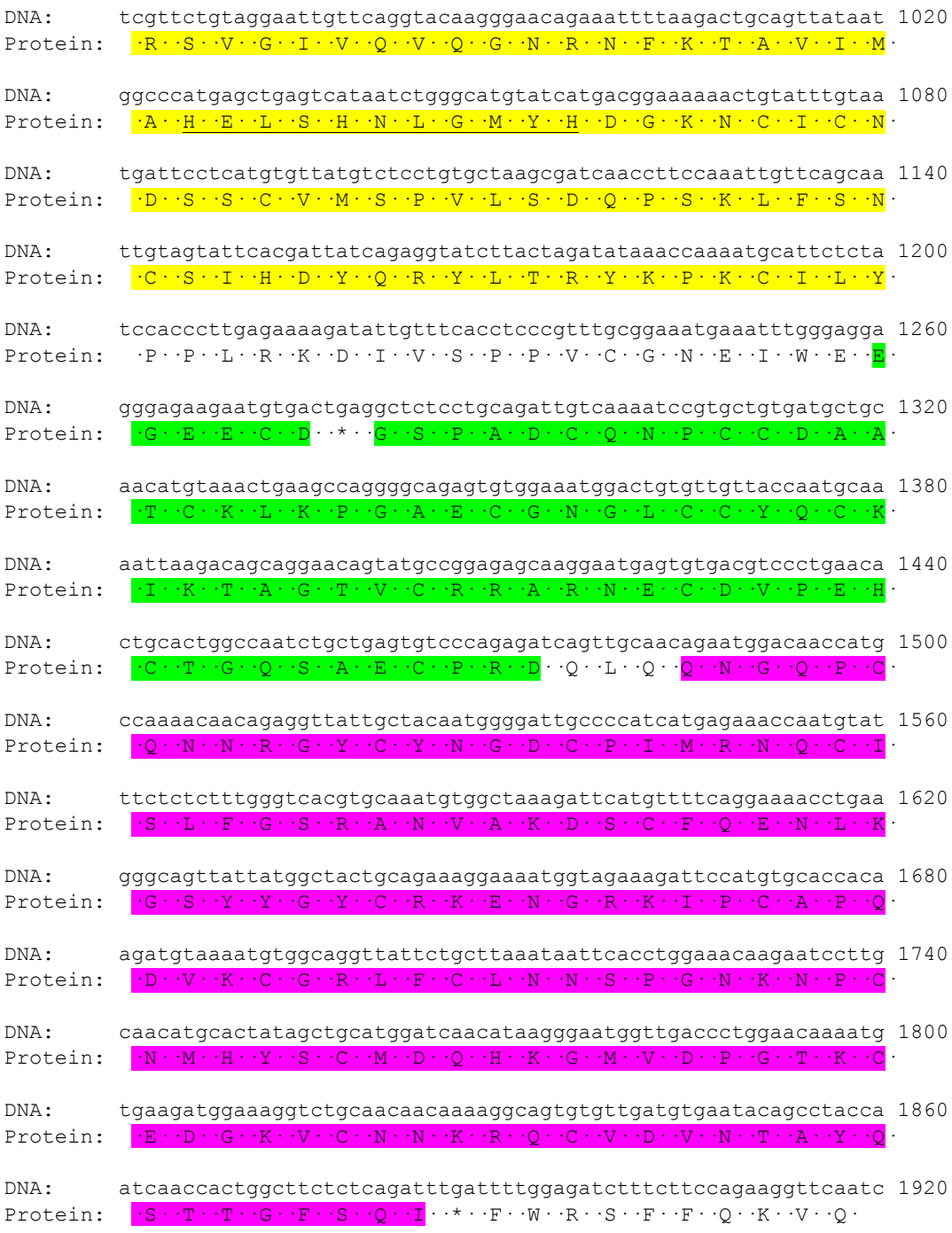


Figure 5.10. Deduced amino acid sequence of RVV-X heavy chain sequence from Sanger sequencing using walking primers. Signal peptide, Cysteine switch motif and zinc binding site are underlined. Prepropeptide region is in blue green. Different domains are highlighted in different colours: Metalloproteinase domain in yellow, disintgerin-like domain in green, cysteine-rich domain in pink.

5.3.3.2. RVV-X light chains

5.3.3.2.1. Primer design and cloning

The primers for RVV-X light chains were designed from the contigs: c20336_g1_i7_M and c20336_g1_i7_M. Three gene specific primers (GSPs) were designed for 5'RACE reaction for 2 RVV-X light chain genes.

Primer		Sequence	nt	Tm	GC %
Gene-specific antisense oligonucleotide (GSP1) / Reverse primer	LC-GSP1	5'-TTGGACCTCCTGACCCATC-3'	19	53.2	58%
Nested gene-specific primer (GSP2)	LC-1-GSP2	5'-ACAGCTGGATCTTAACACTCTGG-3'	23	55.3	48%
	LC-2-GSP2	5'-TCAGGCAGCCTAGAACTTGCACACG-3'	25	61.3	56%
Abridged Anchor Primer (AAP)		5'-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3'	36	92.5	90%

Firstly, 1st strand cDNA was synthesized from total RNA using SuperScript™ II reverse transcriptase with LC-GSP1 which is designed from the common region of the above 2 contigs. The resultant cDNA was tailed with a homopolymeric tail using TdT (Terminal deoxynucleotidyl transferase) and dCTP. PCR amplification was accomplished using Taq DNA polymerase, a nested, gene-specific primer (LC-1-GSP2/LC-2-GSP2) and the Abridged Anchor Primer (AAP).

The conditions for PCR were as follows: initial denaturation for 5 mins at 94°C, followed by 35 extension cycles (denaturation for 1 min at 94°C, annealing for 1 min at 58°C for LC-1 and at 64°C for LC-2, and elongation for 2.5 min at 72°C), and a terminal extension step of 72°C for 10 min (Figure 5.11).

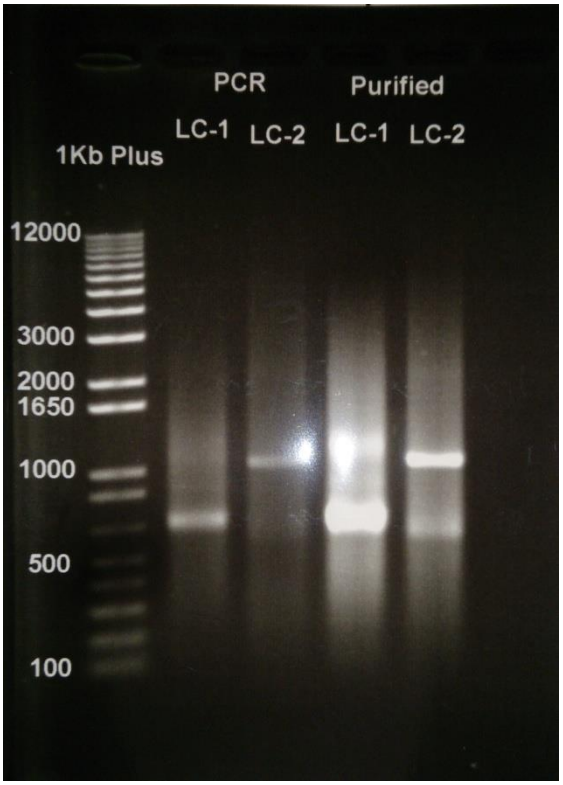


Figure 5.11. PCR products of LC-1-GSP2 and LC-2-GSP2 before and after purification with Wizard PCR Clean-Up System, Promega. Two PCR products, 1200 bp and 600 bp, were found in both samples.

The PCR product was purified using Wizard SV Gel and PCR Clean-Up System (Promega). The purified PCR products were then ligated into pGEM-T Easy vector and transformed into JM109 competent cells, high efficiency (Promega) by a heat-shock method. The transformed cells were cultured on LB/Ampicillin/IPTG/X-Gal agar plates. The white colonies were selected and cloned in LB medium in 37°C shaking incubator at 225 rpm overnight. The plasmids were then purified using GF-1 Plasmid DNA Extraction kit. The cDNA insert was identified on agarose gel (1%) after digested by the BstZI restriction enzyme (Figure 5.12).

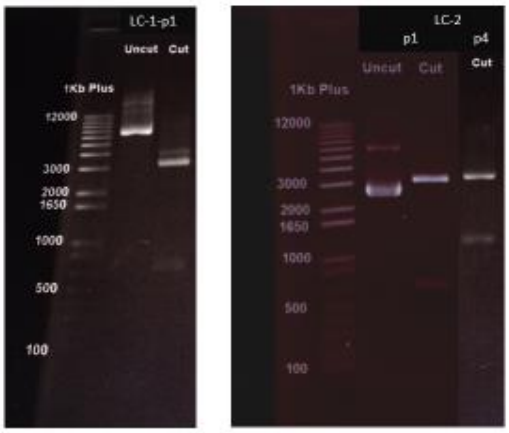


Figure 5.12. Identification of cDNA insert in pGEM-T Easy vector after incubation of plasmids with EcoRI at 37°C for 4 hours. (a) LC-1 plasmid contained 600 bp (LC1-p1) product and LC-2 plasmids contained 1300 bp (LC2-p4) and 600 bp (LC2-p1) products.

5.3.3.2.2. DNA sequencing

The deduced amino acid sequence of LC1-p1 was shown to be identical to that of c200036_g1_i9_M from NGS data and RVV-X light chain 2 from *Daboia siamensis* (ADK22819.1) except for 2 amino acid residues variations: W (tryptophan) to R (arginine) at position 92; and S (serine) to G (glycine) at position 113 (Figure 5.13).

```
RVV_X_LC2           MGRFISVSFGLLVVFLSLSGTGAGLDCPPDSSLYRYFCYRVFKEHKTWEAAERFCMEHPN
c20336_g1_i9_M      MGRFISVSFGLLVVFLSLSGTGAGLDCPPDSSLYRYFCYRVFKEHKTWEAAERFCMEHPN
LC_1_p1             MGRFISVSFGLLVVFLSLSGTGAGLDCPPDSSLYRYFCYRVFKEHKTWEAAERFCMEHPN
                    ************************************************************

RVV_X_LC2           NGHLVSVESMEEAEFVAKLLSNITEKFITHFWIGLMIKDKEQECSSEWSDGSSVSYDNLD
c20336_g1_i9_M      NGHLVSVESMEEAEFVAKLLSNITEKFITHFWIGLMIKDKEQECSSEWSDGSSVSYDNLD
LC_1_p1             NGHLVSVESMEEAEFVAKLLSNITEKFITHFRIGLMIKDKEQECSSEWSDGSGVSYDNLD
                    ****************************** ******************.*******

RVV_X_LC2           KREFRKCFVLEKESGYRMWFNRNCEERYLFVCKVPPEC-(99% coverage & identity)
c20336_g1_i9_M      KREFRKCFVLEKESGYRMWFNRNCEERYLFVCKVPPEC-(99% coverage & identity)
LC_1_p1             KREFRKCFVLEKESGYRMWFNRNCEERYLFVCKVPPEC*
                    ***********************************
```

Figure 5.13. Clustal alignment of translated amino acid sequences from clone LC-1-p1 and c20036_g1_i9_M contig to RVV-X light chain 2 from *Daboia siamensis* (ADK22819.1).

```
c20336_g1_i7_M      MGRFISISFGLLVMFLSLSGTEAVLDCPSGWLSYEQHCYKGFNDLKNWTDAEKFCTEQKK
LC_2_p1             MGRFISVSSGCLVVFLSLSGTEAVLDCPSGWLSYEQHCYKGFNDLKNWTDAEKFCTEQKK
RVV_X_LC1           MGRFISVSFGCLVVFLSLSGTEAVLDCPSGWLSYEQHCYKGFNDLKNWTDAEKFCTEQKK
                    ******:* * **:*********************************************

c20336_g1_i7_M      GSHLVSLHSREEEKFVVNLISENLEYPATWIGLGNMWKDCRMEWSDRGNVKYKALAEESY
LC_2_p1             GSHLVSLHSREEEEFVVNLISENLEYPATWIGLGNMWKDCRMEWSDRGNVKYKALAEESY
RVV_X_LC1           GSHLVSLHSREEEEFVVNLISENLEYPATWIGLGNMWKDCRMEWSDRGNVKYKALAEESY
                    ************:*********************************************

c20336_g1_i7_M      CLIMITHEKVWKSMTCNFIAPVVCKF- (99% coverage, 96% identity)
LC_2_p1             CLIMITHEKEWKSMTCNFIAPVVCKF*
RVV_X_LC1           CLIMITHEKEWKSMTCNFIAPVVCKF- (99% coverage, 99% identity)
                    ********* ***************
```

Figure 5.14. Clustal alignment of translated amino acid sequences from clone LC-2-p4 and c20036_g1_i7_M contig to RVV-X light chain 1 from *Daboia siamensis* (Q4PRD1.2)

The deduced amino acid sequence of LC2-p4 was identical to that of c200036_g1_i7_M from NGS data and RVV-X light chain 1 from *Daboia siamensis* (Q4PRD1.2) except for some amino acid variations (Figure 5.14).

The nucleotide sequence from LC2-p1 did not match to any sequence of the 2 light chain 2 contigs from the transcriptome. ORF finder results showed that it is 335 amino acids long. The Blastx result showed the translated sequence is 90% identity to a hypothetical protein from *Pseudomonas fluorescena* containing PDDEXK_4 domain. This belongs to PD-(D/K)XK nuclease family. It might be a new protein from snake venom or a result of a bacterial contamination.

5.3.4. Analysis of SVMPI transcripts

From the transcriptome of Myanmar Russell's viper venom glands, a total of 4 contigs were annotated as the Snake Venom Metalloproteinase Inhibitors (SVMPI). The conceptually translated proteins were aligned with those transcripts of African vipers, *Echis ocellatus* (A8YPR6) and *Cerastes cerastes cerastes* (A8YPR9). The signal peptides are highly similar and a new tripeptide QRW motif in addition to a QKW motif was found in the MRV transcripts. The tripeptides were flanked by the conserved PXXQ(K/R)WXXP motifs. The SVMPI transcripts of MRV also contained a conserved poly-Gly (pG) motif instead of the poly-His poly-Gly (pHpG) seen in *E. ocellatus* SVMPI transcripts. Moreover, the C-terminal portion of the SVMPI

transcripts of MRV have an atrial natriuretic peptide (ANP) domain in place of the C-type natriuretic peptide (CNP) domain seen in the two African viper SVMPI transcripts (Figure 5.15).

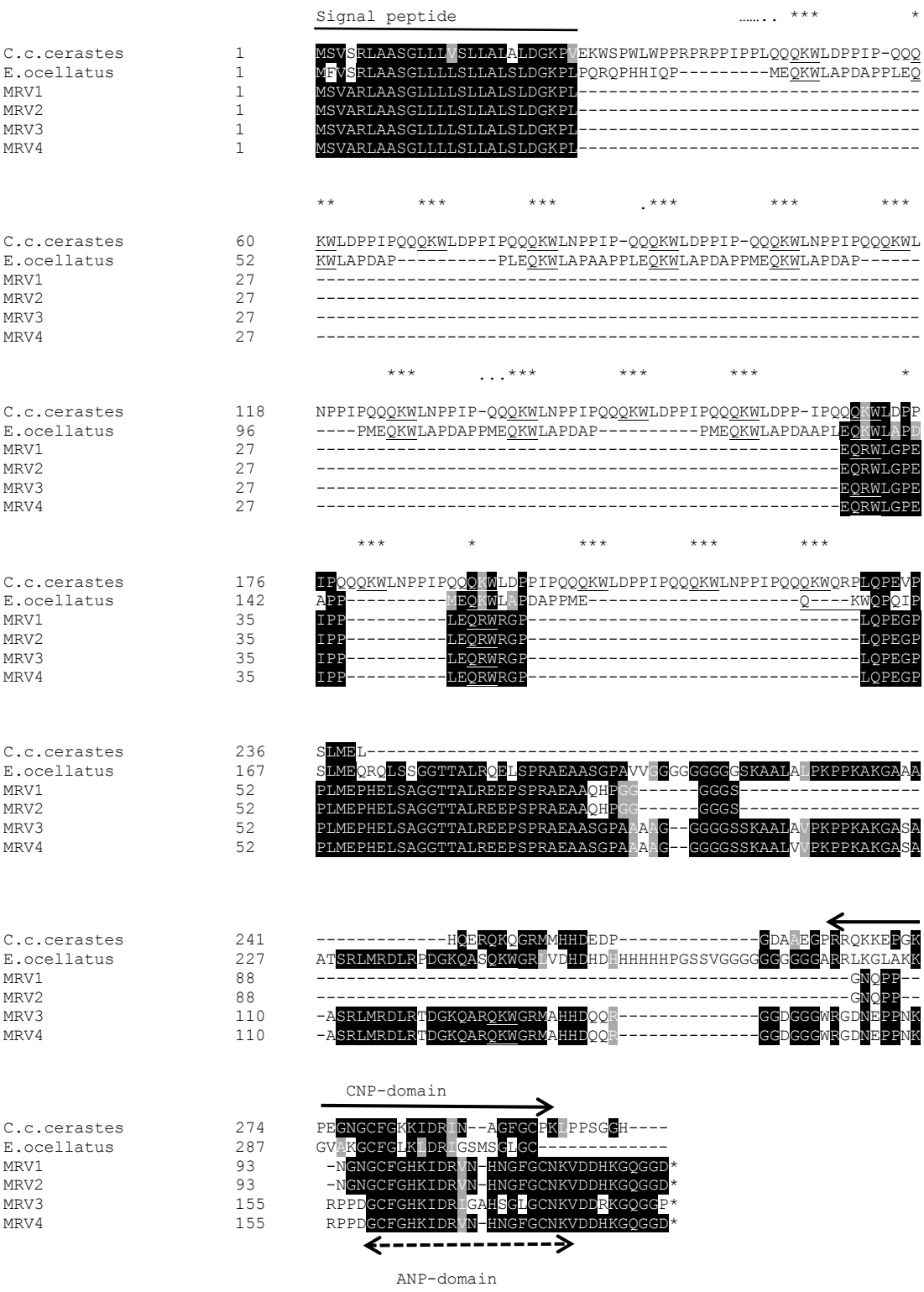


Figure 5.15. Multiple sequence alignment of the polypeptide encoded by Myanmar Russell's viper SVMPI transcripts (MRV1-4) with those of two African vipers [*C. c.*

***cerastes* (A8YPR9) and *E. ocellatus* (A8YPR6)].** The signal peptides are denoted by a solid line above the sequence. The active tripeptides are underlined and identified with three asterisks. The varied residue is defined by a single asterisk. The CNP domains are indicated with a solid arrowed line and ANP domains with a dashed arrowed line.

New SVMPI transcripts from Myanmar Russell's viper were discovered to contain two inhibitory tripeptides, QKW and QRW. The tripeptide sequences are found in the same transcript as natriuretic peptide sequences, as is the case in African vipers. This assortment of different peptide sequences in the same transcript could be related to independent evolution of toxin genes in snakes. The conserved proline residues in the consensus sequence PXXQ(K/R)WXXP might be a signal point for cleavage of tripeptides from the transcripts. The mechanism for release of tripeptides from their transcripts is still unknown. These tripeptides and natriuretic peptides are observed separately in venom, although they are encoded from the same transcript. Since ANP is homologous to a hormone, it might be processed near the effective cells. The release and modification of tripeptides (157) might probably occur during the exocytosis process at an earlier stage than the natriuretic peptides (158).

5.4. Conclusions

In conclusion, the expression profile of SVMPs from Myanmar Russell's viper was analysed to correlate its biological activities with the clinical features of snake-bite patients. The sex-related variation in SVMP transcripts was also identified from the transcriptomic analysis. Analysis of SVMPI transcripts elucidated the potential presence of endogenous tripeptide inhibitors for SVMPs in the snake venom.

The transcriptome of Myanmar Russell's viper venom glands showed that SVMPs were more highly expressed than any other toxin groups. A total of 5 isoforms of disintegrins, 8 isoforms of P-III SVMPs and 2 isoforms of P-II SVMPs were observed in the transcriptomes. The expression profile of SVMP transcripts from MRV demonstrates potential differences in toxicity of male- and female-snake venoms. However, the relative abundance of SVMP transcripts is similar between male and female venom gland transcriptomes. Informatics analysis shows that disintegrin

transcripts are the most highly expressed followed by P-III SVMP and P-II SVMP transcripts. There is a variation in expression of different gene isoforms between analysed male and female glands, which could reflect a sex-dimorphism of viper venom biological activities. The more highly-expressed disintegrin transcripts in female venom gland than in male venom glands might reflect the greater toxicity of female snake venom than male snake venom. The finding of sex-based difference in venom gland transcriptome highlighted that a well-representative venom (i.e., including venoms from different age, sex, geographical region) should be used for animal immunisation in the antivenom production process.

In sequence analysis, disintegrin contains RTS motifs, that selectively block the integrin of basement membrane type IV collagen and laminin-1. Thus, disintegrin from MRV may be responsible for bleeding in patients via weakening of endothelial cells adhesion. The highly-expressed P-III SVMPs in the transcriptomes are RVV-X and VLAIP homologs (here named DSAIP) and both contribute to severe bleeding in patients. RVV-X acts as pro-coagulant to disturb the normal coagulation cascade, resulting massive intravascular coagulopathy. The biological function of DSAIP can be predicted from its deduced amino acid sequence. Like VLAIP, it can inhibit endothelial cell adhesion to the extracellular matrix, thereby resulting in increased vascular permeability. Severe haemorrhage, resulting from intravascular coagulopathy and increased vascular permeability, is a contributing factor for oedema, shock, and renal failure following Russell's viper bites.

There was an interesting finding from the alignment of the RVV-X heavy chain and light chains among different Russell's vipers. The variation in RVV-X light chains, especially in light chain 2, is more obvious among Russell's viper from different countries than in the heavy chain. This finding provides a possible reason for different clinical presentations in different countries. The novel finding of DSAIP suggested that further study on the characterisation of this protein to define its structure and function is needed.

Thus, transcriptome analysis of a toxin gene group shows both qualitative and potentially quantitative (provided the translation of the transcripts is proportional to their abundance) nature of venom composition. The diversity of structural domains

and their role after post-translation modification may also play important role in the presentation of diverse biological activities of a venom toxin group.

Although transcriptomic studies permit the identification of toxin transcripts abundance and their different isoforms, there are some limitations in this approach. For example, post-translation modification of proteins such as glycosylation, oligomerisation which are important for toxin functions cannot be detected in transcriptomic analyses. Additionally, some genetic features such as introns, intergenic regions and cis- and trans-transcriptional regulatory elements that have crucial role in the control of toxin gene expression cannot easily be identified by transcriptome studies. Moreover, the actual molecular composition of venom needs to be elucidated by a combined proteomics and transcriptomics approach, to ensure that transcripts are actually translated into functional proteins.

In this thesis, only the SVMP toxins group was analysed in detail, but there are other important toxins in our venom gland transcriptome that should be analysed in detail in future studies. It was not possible to fully analyse all of these transcripts in the timecourse of this project. Since the genome of Russell's viper has not been sequenced, the manual annotation of toxins against ESTs from NCBI databases has some limitations. The annotation of toxin transcripts could not be confirmed only with Blastn and Blastx searches. It required detail analysis of sequences one by one to identify the signal peptide and functionally-conserved domains of toxins with the signal peptide prediction tool (SignalP 4.1 server) and the NCBI Conserved Domain search tool, respectively. Functional annotation in non-model organisms is mostly based on the homology to annotated genes to a model organism and many toxin genes were called from genes of non-venomous function. Therefore, transcripts might be annotated with the wrong function and manual annotation is needed to counter-check with conserved domain searches and detailed sequence analysis. In a study by Castoe, et al (2014) (159), a total of 25,385 genes was identified from the transcriptome of Burmese python after mapping to its own genome. In this thesis, using a non-model species, the total of toxin genes/transcripts was only predicted from annotation of results Blastn and Blastp, in combination of results from conserved domain searches.

The cloning result of SVMPs showed that the presence of truncated RVV-X protein in snake venom was due to a point mutation at the disintegrin-like domain. The RVV-X and disintegrin might be derived from the same ancestral gene, but this would need to be confirmed by analysis of their genomic sequences. The sequence alignment of RVV-X light chains from NGS and Sanger sequencing after cloning showed 96-99% similarity with 99% coverage, supporting that the result from NGS is compatible with those from Sanger sequencing.

For therapeutic purposes, novel SVMP inhibitory peptides can be predicted from transcripts analysis of toxin inhibitors. These tripeptides were purified and identified from the MRV venom in next Chapter. The finding of a C-type natriuretic peptide (CNP) in SVMPI transcripts needs further study of their medical importance in future.

Chapter 6. Snake venom metalloproteinases and tripeptides from Myanmar Russell's viper venom

6.1. Introduction

Russell's viper (*Daboia russelii*) is a medically important snake, variants of which are distributed throughout East and Southeast Asia. The cause of death includes shock, massive bleeding and renal failure. Snake venom metalloproteinases (SVMPs) play a major role in the local and systemic clinical manifestations: blistering, necrosis and bleeding from the fang marks and incoagulable blood, thrombocytopenia, spontaneous systemic bleeding, hypotension, increased permeability and reduced urine output (8). Although Russell's viper antivenoms are available, their efficacy in reversal tissue damage, such as acute renal failure, is limited (160). Novel treatment modalities are required.

In order to protect against auto-digestion by SVMPs, snake venom of several species are found to contain natural protease inhibitors: citrate and small peptides. The latter bind selectively to SVMPs in the venom glands to protect glandular tissues and venom factors from self-digestion by SVMPs (106). Three endogenous peptides: pyroGlu-Lys-Trp (pEKW), pyroGlu-Asn-Trp (pENW) and pyroGlu-Gln-Trp (pEQW) isolated from venom of Taiwan habu (*Trimeresurus mucrosquamatus*) showed an inhibitory action on proteolytic activity of metalloproteinases present in the crude venom (107). It is reported that these peptide inhibitors regulate the proteolytic activities of their SVMPs in a reversible manner under physiological conditions (108). Other pit vipers, such as *Bothrops asper* (109) and some rattlesnakes (110), also have venoms containing endogenous tripeptides: pEQW and pENW. African vipers, *Echis ocellatus* and *Cerastes cerastes cerastes,* have pEKW tripeptides. These tripeptides are encoded by tandemly repeating elements from the transcripts which also contain a CNP (C-type natriuretic peptide) homologous sequence at the C-terminus (111). Two peptides: PtA (pENW) and PtB (pEQW) isolated from venom liquor of *Deinagkistrodon acutus* (Hundred-pacer viper) showed anti-human platelet aggregation activity in vitro and protection effects on ADP-induced paralysis and formation of pulmonary thrombosis in mice(161).

6.2. Aims

To date, the only treatment for snake-bite is the antivenom which is raised from the injection of a small amount of snake venom into horses or sheep. There is a risk of anaphylactic reactions in patients due to albumin, complement and proteins from the serum of immunised animals. Moreover, the efficacy of the conventional antivenom is low for small toxins, which may escape from recognition by the immune system of hyper-immunised animals and/or a dilution effect of redundant antibodies to large non-toxin venom antigens. Thus, new modality of treatments to neutralize the whole venom effectively are needed.

The goal of this Chapter was to study the effect of SVMPI tripeptides on major SVMPs to possible therapeutic usage of those tripeptides in snake-bite treatment. The objectives were: 1) to purify and identify major SVMPs and specific SVMP inhibitors (SVMPIs) from the venom; and 2) to study the inhibitory actions of synthetic tripeptides on purified SVMPs.

6.3. Results and discussion

6.3.1. Purification and identification of SVMPs

6.3.1.1. Purification of SVMPs

The crude venom of Myanmar Russell's viper (MRV) was initially separated on a Superdex 200 column. Of the three major protein-containing peaks, only the first possessed caseinolytic activity (Figure 6.1). These fractions were pooled and further purified on a Resource Q anion-exchange column. The proteins resolved into two peaks and the first peak (Q1) exhibited caseinolytic activity (Figure 6.2a). The purity of proteins in Q1 was determined on both reducing and non-reducing SDS-PAGE. Non-reducing SDS-PAGE of this fraction showed it to contain two bands at 85 kDa and 67 kDa. Under reducing conditions, the main protein bands ran at approximately 67 kDa band and low molecular weight (15-20 kDa) bands were evident.

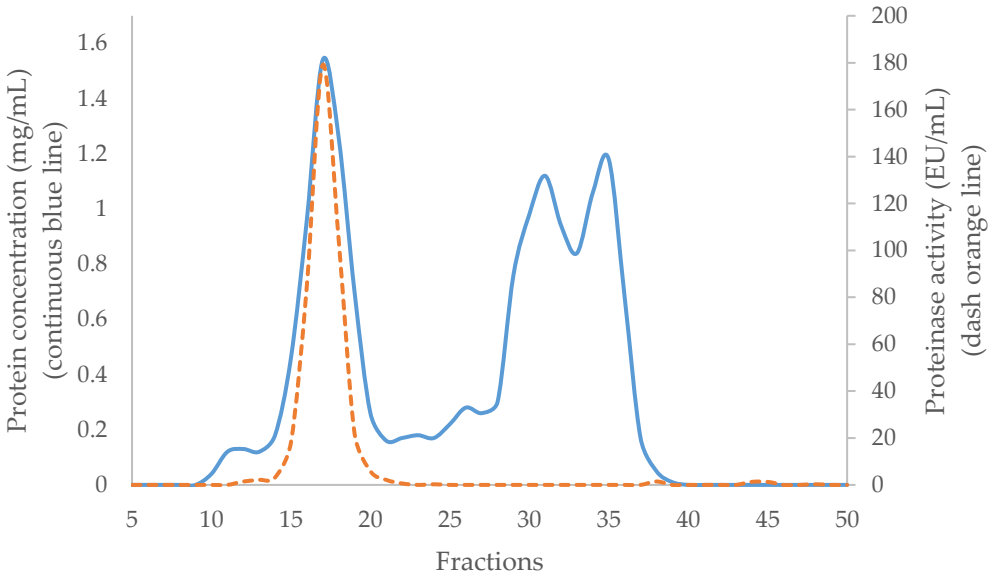
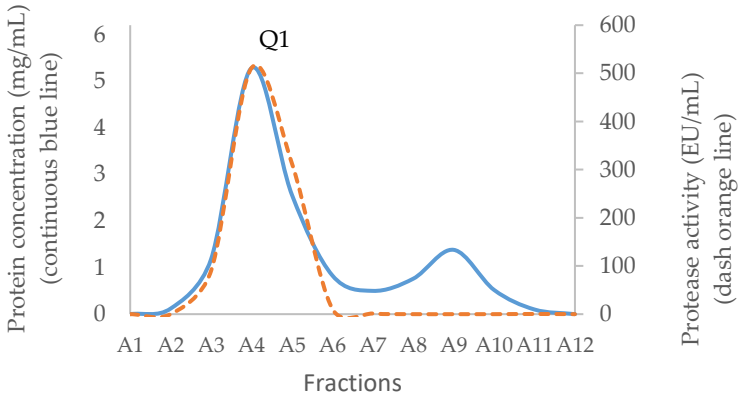


Figure 6.1. Fractionation of Myanmar Russell's viper crude venom through Superdex 200 gel filtration column. Crude venom was separated in 0.01 M phosphate buffered saline (pH 7.4) at 2 mL/min. Each fraction was 6 mL in volume. The blue continuous line shows the protein concentration (mg/mL) and the orange dashed line shows protease activity (EU/mL) in collection fractions.

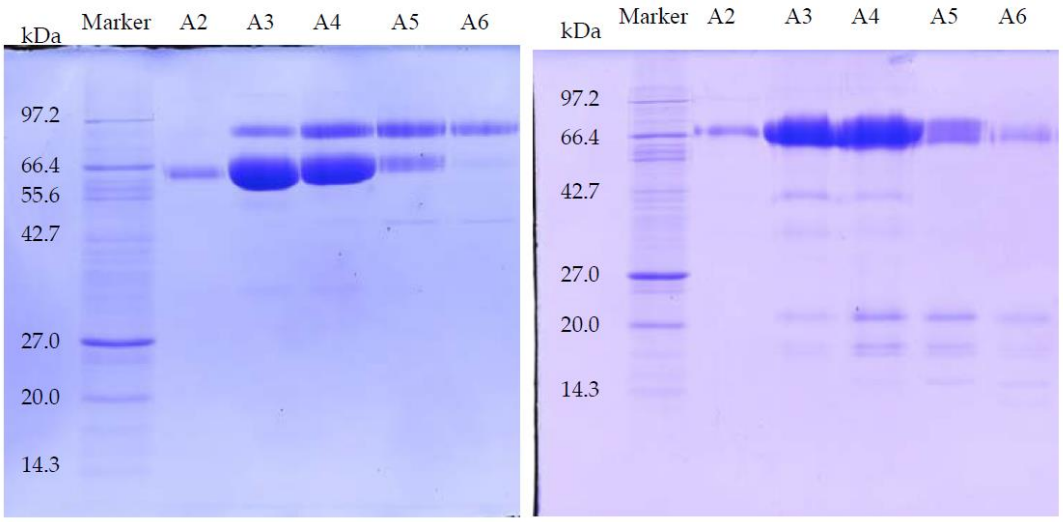
Table 6.1. The protein concentration (mg/mL) and protease activity (EU/mL) in purification fractions after Superdex 200 gel column chromatography

Tube	Protein concentration (mg/mL)	Protease activity (EU/mL)	Tube	Protein concentration (mg/mL)	Protease activity (EU/mL)	Tube	Protein concentration (mg/mL)	Protease activity (EU/mL)
8	0	0	20	0.26	5.88	32	0.94	0
9	0	0	21	0.16	2.14	33	0.84	0
10	0.04	0	22	0.17	0.61	34	1.06	0
11	0.12	0	23	0.18	0	35	1.18	0
12	0.13	1.46	24	0.17	0.27	36	0.68	0
13	0.12	2.14	25	0.22	0	37	0.17	0
14	0.18	3.16	26	0.28	0	38	0.05	1.46
15	0.46	18.12	27	0.26	0	39	0.01	0
16	0.96	84.76	28	0.30	0	40	0	0
17	1.54	179.11	29	0.76	0	41	0	0
18	1.26	104.76	30	0.98	0	42	0	0
19	0.70	22.03	31	1.12	0	43	0	0

This material (Q1) was then subjected to further separation on either HIC for activity studies, or RP-HPLC when proteins were prepared for mass spectrometry. A Phenyl Superose column was used for HIC during which the protein fraction resolved into 2 peaks: H1 (eluted at 13 minutes), and H2 (eluted at 29 minutes), respectively (Figure 6.3a-c). For RP-HPLC, a Phenomenex Luna C4 column was used and again the proteins were separated into 2 peaks (R1 and R2) (Figure 6.3d-f). SDS-PAGE analysis and activity studies showed H1 to be the same protein as R1 running at 85 kDa under non-reducing conditions, but at 67 kDa with several subunits at 15-20 kDa when reduced. H2 is the same as R2, with a single band at 68 kDa under both reducing and non-reducing conditions.



(a)



(b) (c)

Figure 6.2. Separation of fractions 15-18 from GFC on a Resource Q anion-exchange column. (a) Chromatography trace showing protein concentration and caseinolytic activity. Peak one (Q1) contained fractions with protease activity. SDS-PAGE of the purified proteins (Fraction A2-A6) under (b) non-reducing and (c) reducing conditions.

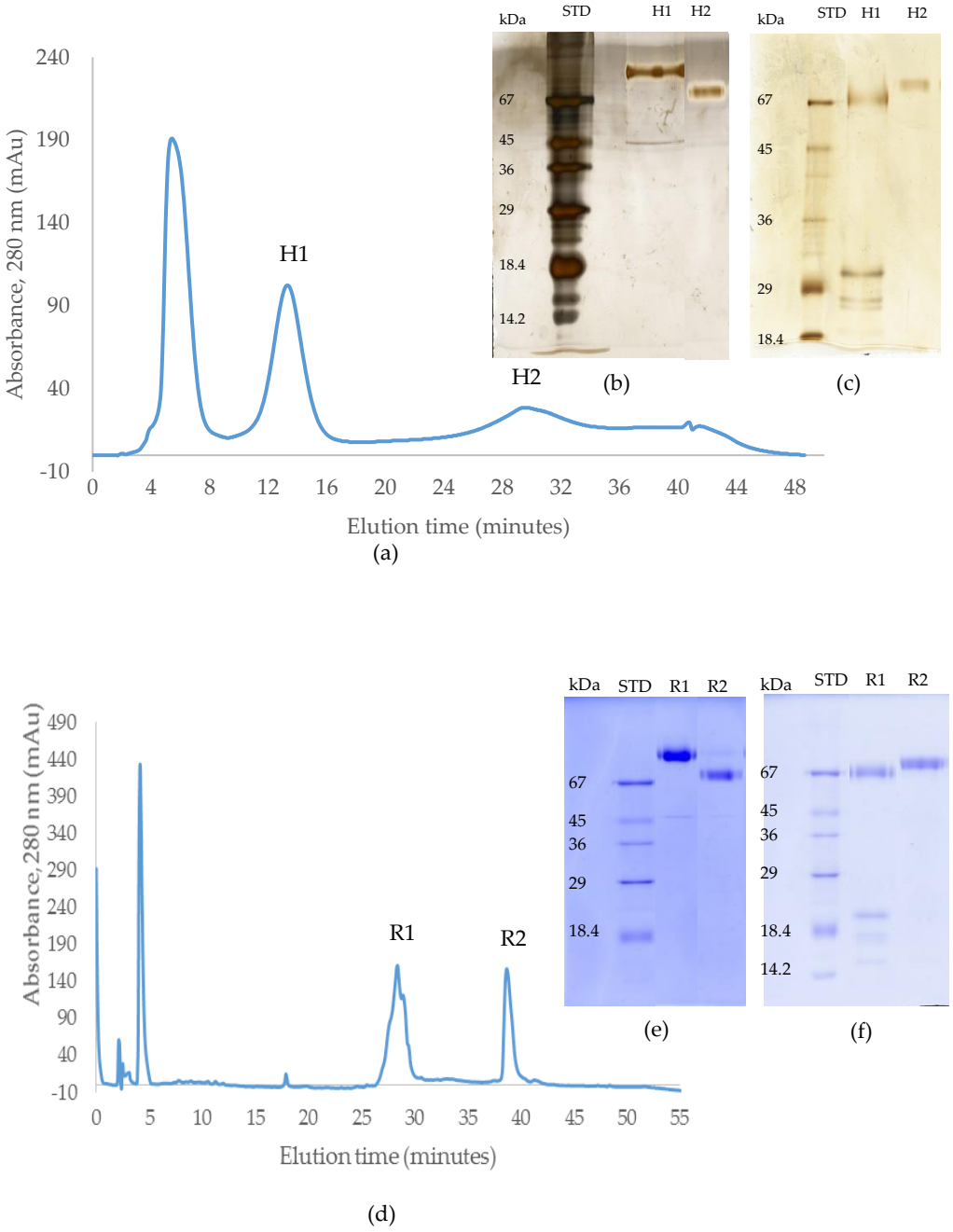


Figure 6.3. Further purification of fraction A2-A6 from Resource Q column. (a) Chromatography of fraction Q1 from the Resource Q column on Phenyl Superose column (HIC) showing protein-containing peaks (H1 and H2); (b) non-reducing and (c) reducing SDS-PAGE of purified proteins (silver-stained). (d) Chromatography of fraction Q1 from Resource Q column on a C4 RP-HPLC column. Two protein peaks were observed: R1 and R2; SDS-PAGE of purified proteins R1 and R2 (e) under non-reducing conditions and (f) reducing conditions. The early peaks in (a) and (d) contained no protein when analysed by SDS-PAGE.

6.3.1.2. Identification of SVMPs

Both proteins with protease activity were purified by C4 RP-HPLC in preparation for mass spectrometric analysis [see R1 and R2 in Figure 6.3d]. For R1, the protein was reduced and treated with iodoacetamide and digested with trypsin in the presence of 2 M urea and the digest was analysed using LC-ESI-MS/MS (Figure 6.4). LC-ESI-MS/MS analysis of the tryptic peptides provided sufficient sequence coverage to match to the mature sequence (residues 189-615) of the Eastern Russell's viper (*Daboia russelii siamensis*) RVV-X H chain VM3CX_DABSI (Q7LZ61) (Figure 6.4a). In the same digest mixture we also found matches to RVV-X light chain proteins LC1 SLLC1_DABSI (Q4PRD1) (Figure 6.4b) and LC2 SLLC2_DABSI (Q4PRD2) (Figure 6.4c) from the same species.

The protein R2 from RP-HPLC was digested in the same way as R1. In this case MALDI-MS analysis was used to identify tryptic peptides that matched the mass [$M+H^+$] of those predicted from the sequence of Daborhagin-K (Indian Russell's viper) (VM3DK_DABRR) (B8K1W0) (Figure 6.5). The majority of the most abundant peptides matched the mass of expected tryptic peptides, and notably we found many of the same tryptic peptides as did Chen et al. [Table 2 in ref. 162] for Daborhagin-M.

As a result of this work we can identify R1, H1 as the Myanmar Russell's viper RVV-X and R2, H2 as Myanmar Russell's viper Daborhagin and will to refer them as such from hereon.

1 *MMQVLLVTISLAVFPYQGSS*IILESGNVNDYEVVYPQKVTALPKGAVQQPEQKYEDTMQY 60

61 EFEVNGEPVVLHLEKNKILFSEDYSETHYYPDGREITTNPPVEDHCYYHGRIQNDAHSSA 120

121 SISACNGLKGHFKLRGEMYFIEPLKLSNSEAHAVYKYENIEKEDEIPKMCGVTQTNWESD 180

181 KPIKKASQLVSTSAQFNKIFIELVIIVDHSMAKKCNSTATNTKIYEIVNSANEIFNPLNI 240

241 HVTLIGVEFWCDRDLINVTSSADETLNSFGEWRASDLMTRKSHDNALLFTDMRFDLNTLG 300

301 ITFLAGMCQAYRSVEIVQEQGNRNFKTAVIMAHELSHNLGMYHDGKNCICNDSSCVMSPV 360

361 LSDQPSKLFSNCSIHDYQRYLTRYKPKCIFNPPLRKDIVSPPVCGNEIWEEGEECDCGSP 420

421 ANCQNPCCDAATCKLKPGAECGNGLCCYQCKIKTAGTVCRRARDECDVPEHCTGQSAECP 480

481 RDQLQQNGKPCQNNRGYCYNGDCPIMRNQCISLFGSRANVAKDSCFQENLKGSYYGYCRK 540

541 ENGRKIPCAPQDVKCGRLFCLNNSPRNKNPCNMHYSCMDQHKGMVDPGTKCEDGKVCNNK 600

601 RQCVDVNTAYQSTTGFSQI 619

(a)

MGRFISVSFGCLVVFLSLSGTE GVLDCPSGWLSYEQHCYK
GFNDLKNWTDAEKFCTEQKKGSHLVSLHSREEEEFVVNLISENLEYPATWIGLGNMWKDCRM
EWSDRGNV KYKALAEESYCLIMITHEKEWKSMTCNFIAPVVCKF

(b)

MGRFISVSFGLLVVFLSLSGTGA GLDCPPDSSLYRYFCYRVFKEHKTWEA
AERFCMEHPNNGHLVSIESMEEAEFVAKLLSNTTGKFITHFWIGLMIKDK
EQECSSEWSDGSSVSYDKLGKEEFRKCFVLEKESGYRMWFNRNCEERYVFVCKVPPEC

(c)

Figure 6.4. Data from LC-ESI-MS/MS analysis of the tryptic peptides from purified RVV-X. The data was obtained by digesting the R1 fraction from RP-HPLC with trypsin. The prepro-sequences of (a) RVV-X H chain (VM3CX_DABSI; Q7LZ61), (b) RVV-X light chain 1 [LC1 SLLC1_DABSI (Q4PRD1), 14.5 kDa], and (c) RVV-X light chain 2 [SLLC2_DABSI (Q4PRD2), 15.9 kDa] annotated to show the peptides (underlined) identified in this analysis. The signal peptides are in italic.

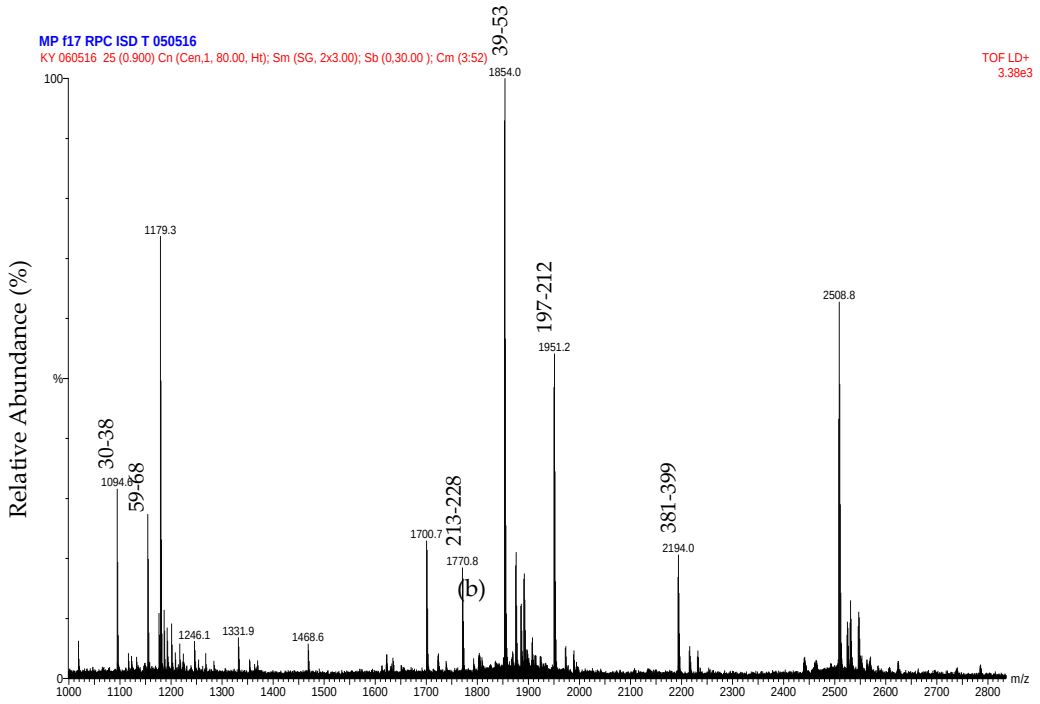


Figure 6.5. MALDI-MS spectrum of tryptic peptides from purified daborhagin. The data was obtained by digesting the R2 fraction from RP-HPLC with trypsin. The numbers above the peptides masses indicates the residue numbers for the peptides matched to the sequence of Daborhagin K (VM3DK_DABRR) (B8K1W0). All m/z values are for the $M+H^+$ ions. The ions at 1854 and greater have been labeled with the m/z value for the ion containing one carbon as the C^{13}isotope.

Russell's viper is a venomous species of the South-East Asian region. The clinical manifestations of its bites reflect the high content of proteases such as snake venom serine proteases and snake venom metalloproteinases (SVMPs). It has been shown that the SVMPs comprise approximately 11% to 65% of the total protein in the Viperidae venoms (76). In Myanmar Russell's viper, SVMPs contribute to 20% of the crude venom (data not shown) and Class III SVMPs are found to be the major component. In comparison with other species, the Myanmar species have 6-7 times more daborhagin than Indian species (162) and SVMPs, mainly RVV-X, in Sri Lankan species comprise just 6.9% of the crude venom (20). Thus, it can be noted that the

Myanmar venom contains greater amounts of SVMPs than that of the Indian and Sri Lankan species. The variations in types and amounts of SVMPs in venom among different subspecies of Russell's viper might be due to diversity in their prey at different locations and this could lead to the dissimilar severity or clinical presentations of snakebite patients.

Russell's viper venom factor X activator (RVV-X) is a well-characterized Class III metalloproteinase (formally known as Class IV) which specifically activates coagulation factor X by hydrolysis of an Arg-Ile bond in factor X. It is a glycoprotein consisting of a heavy chain (α-chain, 57.6 kDa) and two light chains (β- and ϒ-chains, 19.4 kDa and 16.4 kDa) linked by disulfide bonds (96). In addition to proteolytic activity on factor X and IX, RVV-X also inhibits collagen- and ADP-stimulated platelet aggregation (29) and has a strong affinity for protein S (97). Factor X activators are also found in *Vipera lebetina* (blunt-nosed viper) in which it exhibits specific proteolytic activity towards human factor X and also factor IX, but it is not active against prothrombin nor fibrinogen (163) .

In the present study, the purified RVV-X was shown to be composed of a heavy chain (67 kDa) and two light chains (20 kDa and 15 kDa). The two bands on SDS-PAGE at around 15 kDa level suggested that the γ-light chain in Myanmar species might exist as 2 forms, likely due to either amino acid variation or differences in N-glycosylation. Our experiments showed that MRV RVV-X possesses hydrolytic activity to gelatin (Type I collagen, bovine), which had not been characterized before for RVV-X.

Another potent Class III SVMP, Daborhagin, composed of metalloproteinase, disintegrin and cysteine-rich domains, was also purified from MRV venom. The Daborhagin-M from Myanmar Russell's viper venom specifically digested the α-chain of fibrinogen, fibronectin and type IV collagen in vitro and exhibited haemorrhagic (34), oedema inducing and myonecrotic activity in mice (26). In our studies, a 67 kDa metalloproteinase was isolated and matched to Daborhagin-K from Indian species using mass spectrometric analysis of tryptic peptides. This MRV Daborhagin exhibited potent α-fibrinogenolytic activity, but did not digest gelatin.

In the current purification strategy, the two SVMPs were co-purified initially, but then could be separated from each other using either hydrophobic interaction chromatography or RP-HPLC. Better resolution was evident on RP-HPLC, and the presence of multiple forms of RVV-X was indicated by the irregularity of the RVV-X RP-HPLC peak, suggesting heterogeneity of the protein (R1, Figure 6.3d). Two isoforms of the heavy chain and 6 isoforms of the light chain from RVV-X have been revealed on 2-D electrophoresis in the proteomic study of Risch, et al in the same species (146).

6.3.2. Purification and identification of tripeptides

The low molecular fractions from Superdex 200 chromatography were analysed using C18 RP-HPLC. Fraction 48 was found to contain the highest concentration of the tripeptides. Upon RP-HPLC analysis of this fraction, two peaks (A_p and B_p) eluted close together at 31-33 minutes (Figure 6.6). These peaks possessed the same elution time as that of two synthetic peptides pEKW (peak A_s) and pERW (peak B_s), respectively. RP-HPLC analysis of mixtures of natural and synthetic tripeptides showed perfect co-chromatography. The purified endogenous tripeptides were then analyzed using ESI-MS. The resultant spectra of peak A_p showed a strong $M+H^+$ ion at m/z 444.2, (the predicted monoisotopic mass of pEKW is 443.2). Analysis of peak B_p, also showed a strong $M+H^+$ ion at m/z 472.2(the predicted monoisotopic mass of pEKW is 471.2) (Figure 6.7a,c). MS/MS analysis of these tripeptides produced a set of fragment ions consistent with their expected amino acid sequence (Figure 6.7b,d).

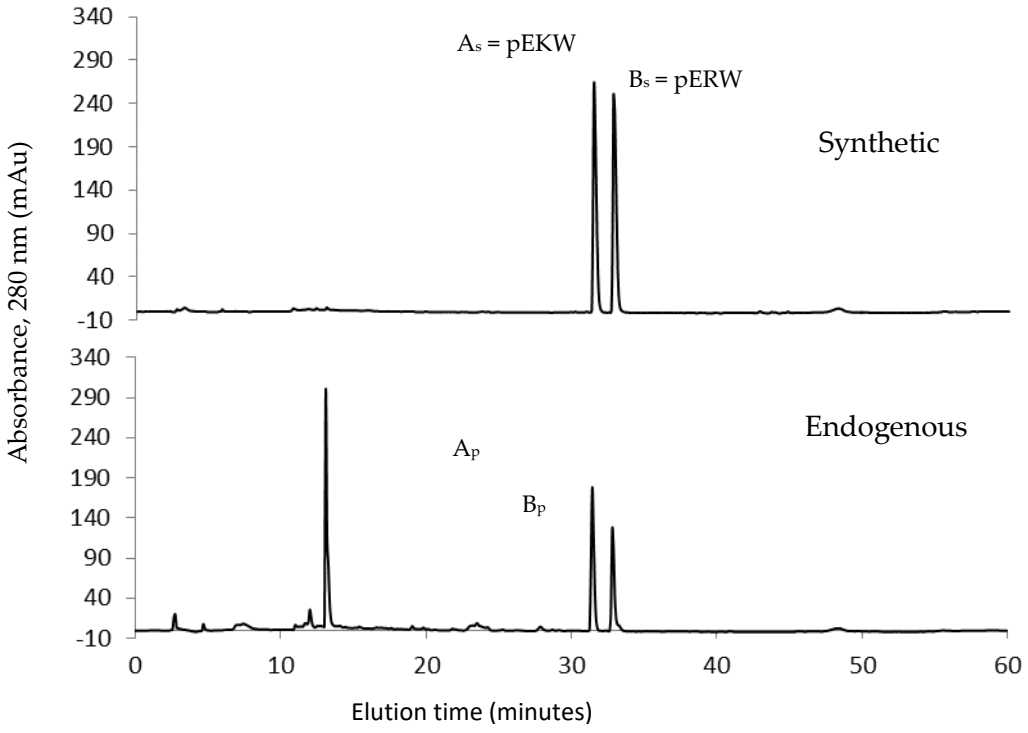
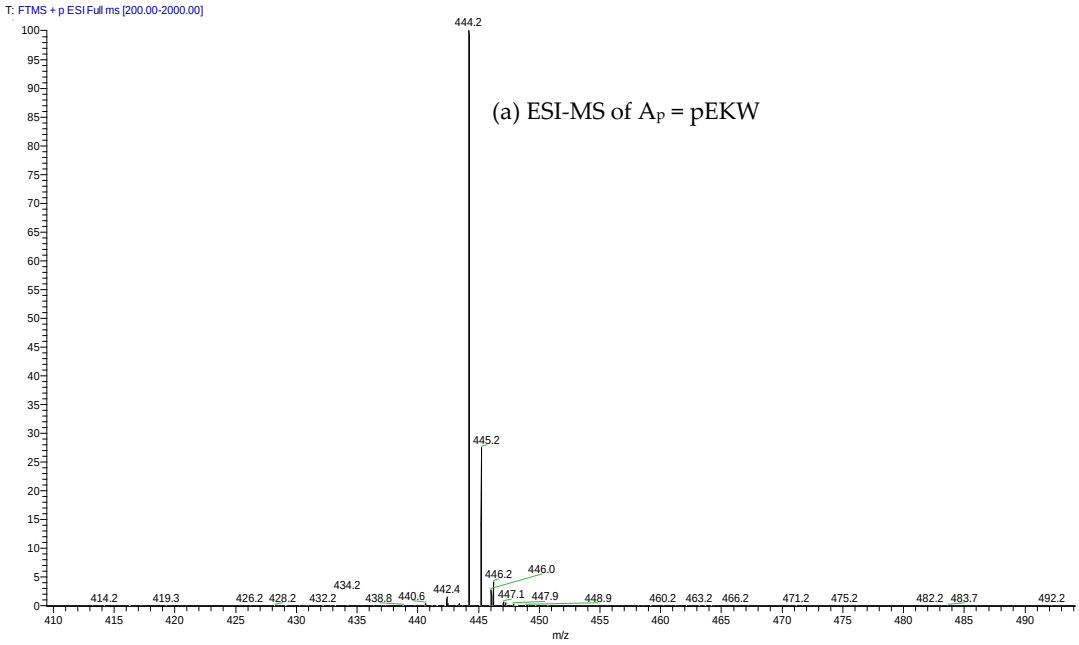
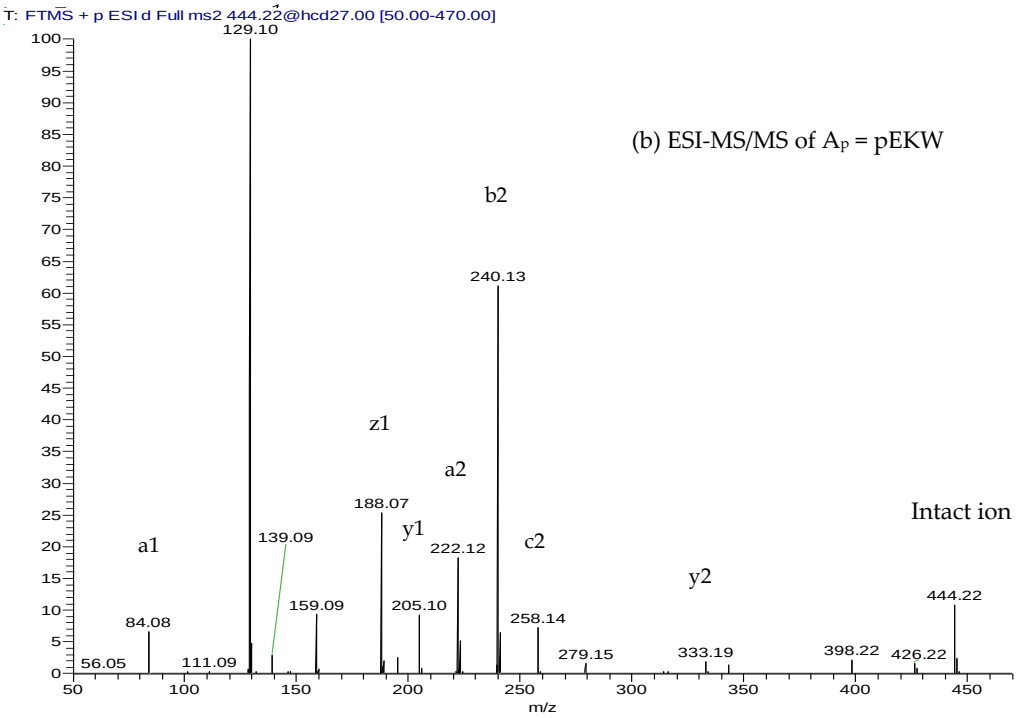
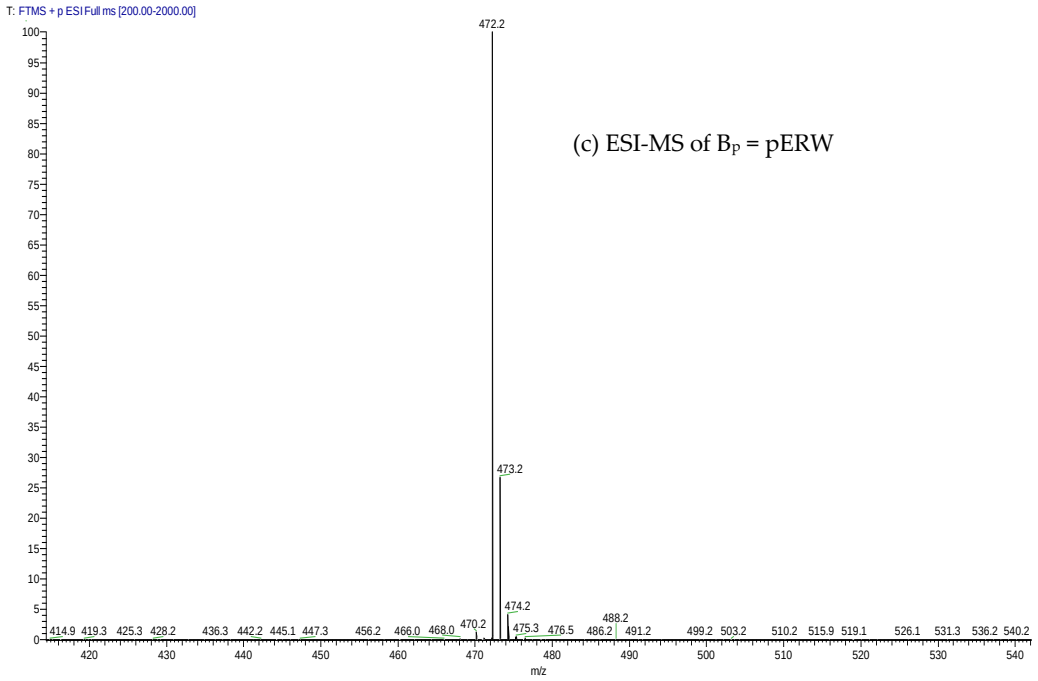


Figure 6.6. C18 RP-HPLC analysis of tripeptides. Upper panel: synthetic pEKW and pERW (1µg of each). Lower panel fraction 48 (200 µL) obtained from gel filtration chromatography of crude venom. Peak A_s and peak B_s represent the two synthetic tripeptides. Peak A_p and peak B_p represent the two tripeptides from Fraction 48.





(b) ESI-MS/MS of A_p = pEKW



(c) ESI-MS of B_p = pERW

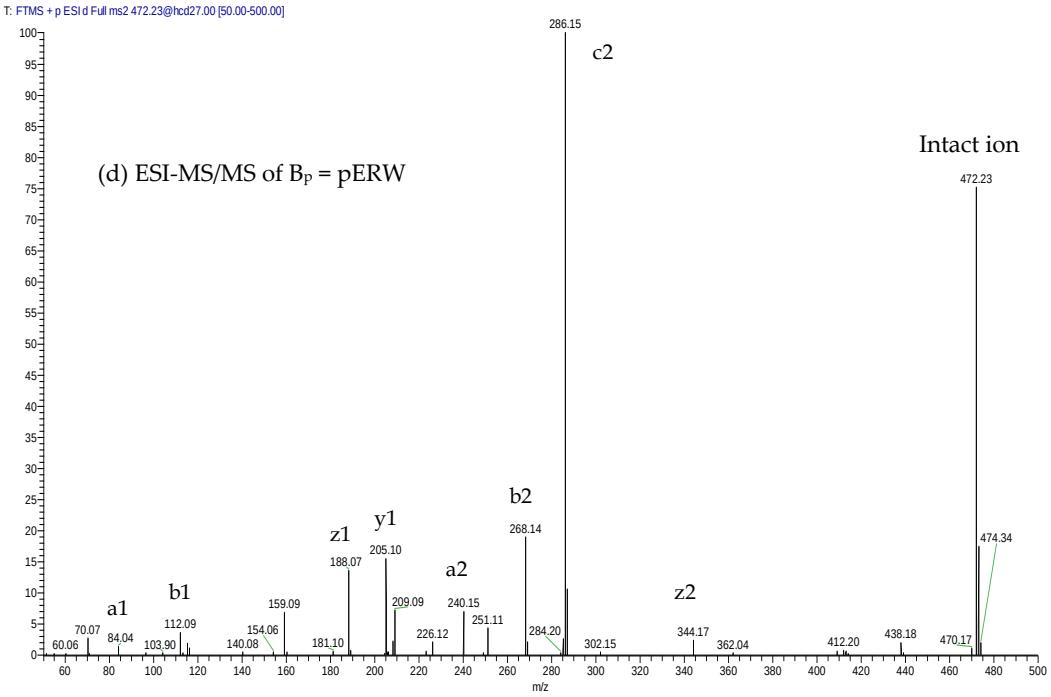


Figure 6.7. ESI-MS and ESI-MS/MS spectra of peak A_p and peak B_p isolated via RP-HPLC of low molecular material obtained from GFC of crude MRV venom. (a, c) ESI-MS spectra. The values indicated are for the $M+H^+$ ions. These are within 0.05 Da of the predicted values for pEKW and pERW (monoisotopic masses are 443.2 and 471.2 respectively). (b, d) ESI-MS/MS spectra. The predicted a, b, c, y and z ions are indicated above the mass values.

The inhibitory tripeptides purified from the MRV venom are in pyroglutamate forms, pEKW and pERW. These were identified using RP-HPLC (co-chromatography with synthetic peptides), LC-ESI-MS analysis of intact mass and LC-ESI-MS/MS sequencing. Although the tripeptide pEKW purified from MRV venom has been found in other snake species, such as *Trimeresurus mucrosquamatus* (107), *Echis ocellatus*, and *Cerastes cerastes cerastes* (111), the tripeptide pERW purified here has not been found in the venom of any other snake species.

6.3.3. Characterization of RVV-X and daborhagin

6.3.3.1. Effect of pH on caseinolytic activity of purified SVMPs

The purified fractions A4 and A5 from Resource Q chromatography (i.e. total SVMPs) were used to determine the working pH range for the enzymes and the effect of metal chelators and citrate on their caseinolytic activity.

The caseinolytic activity of purified enzyme was determined at different pHs, ranging from pH 5 to pH 10, with a substrate concentration of 20 mg/ml. The activity was determined by the amount of tyrosine released due to protease action. The enzyme shows optimal caseinolytic activity at pH 8 (Figure 6.8).

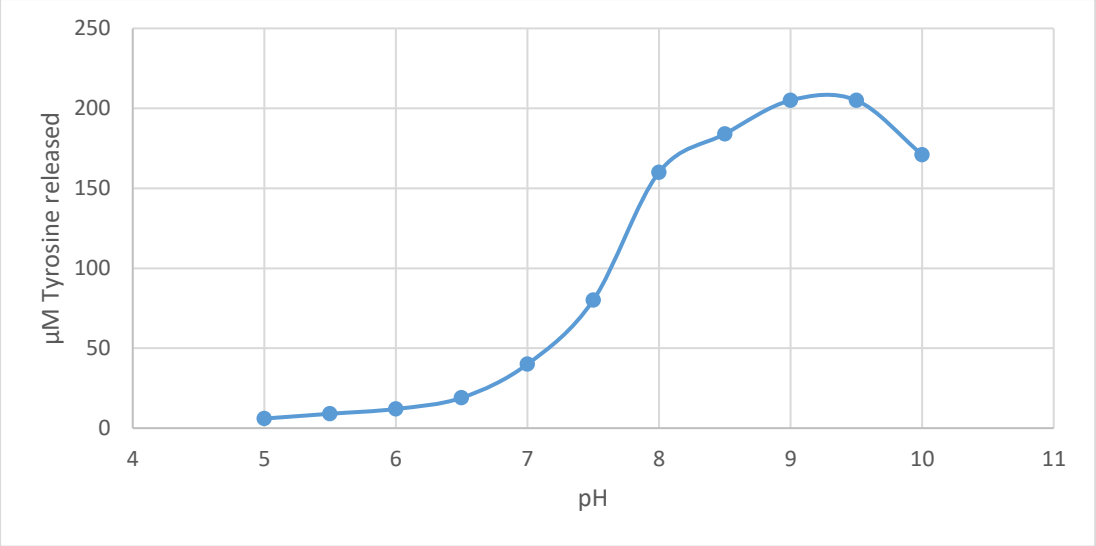


Figure 6.8. Effect of pH on caseinolytic activity of purified fractions of total SVMPs.

6.3.3.2. Inhibition of purified SVMPs by metal chelators

The SVMPs (total 20 µg; approx. 570 at 0.04 mg/ml) were incubated with metal chelators (EDTA, EGTA, or 1, 10-phenanthroline) for 10 min at 37˚C and then assayed for residual protease activity (118) . For the assay with added citrate, 0.1 µg; approx. 3 pmol (final concentration 0.05 mg/ml) SVMP was used and the reaction volume was reduced from 500 µl to 50 µl.

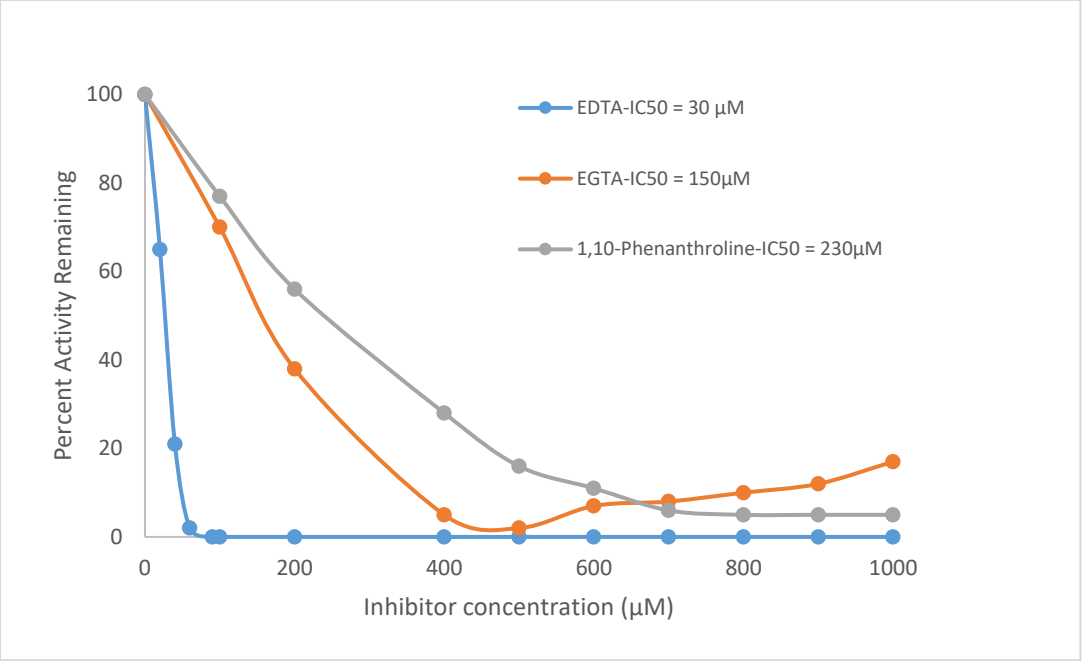


Figure 6.9. Inhibitory effect of metal chelators on caseinolytic activity of purified fractions of total SVMPs.

When compared with the other chelators, EDTA showed the highest inhibitory efficacy against the purified metalloproteinases, the IC_{50} was 30 µM significantly lower than that for EGTA (IC_{50}=150 µM) and 1, 10-phenanthroline (IC_{50}=230 µM) (Figure 6.9). This might be due to the strong stabilization of Zn ion with chemical bonds distributed by EDTA molecule. Although proteinase was completely abolished at inhibitor concentration above 100 µM for EDTA or 1 mM for 1, 10-phenanthroline, it was raised again at inhibitor concentrations above 500 µM for EGTA. This sensitivity to the metal chelators (EDTA, EGTA and 1, 10-phenanthroline) confirmed that the purified enzymes were metalloproteinases.

6.3.3.3. Inhibition of purified SVMPs by citrate

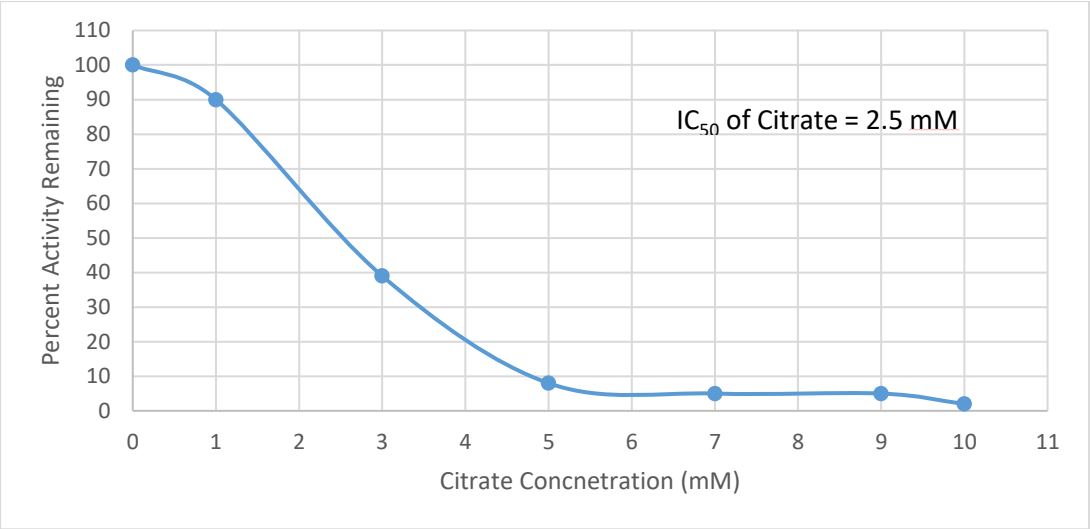


Figure 6.10. Inhibitory effect of citrate on caseinolytic activity of purified fractions of total SVMPs.

Citrate is found to be present in crude venom, where it may inhibit their activity. The percent protease activity remaining is nearly zero at a 5 mM concentration of citrate. The IC_{50} for citrate was found to be 2.5 mM (Figure 6.10). In comparison with the inhibitory action of citrate on CVO protease V, the dominant metalloprotease from *Crotalus viridis oreganus*, more than 65% of enzyme activity is retained at 100 mM citrate concentration (164). Although citrate is not effective in inhibition of protease activity of the metalloproteinase in above species, it is effective to metalloproteinases in our species. Thus, citrate may play an important role in the temporary inactivation of MRV SVMPs during storage in the venom gland.

6.3.3.4. Gelatinolytic activity

The purified proteins RVV-X and Daborhagin from HIC were used for characterization of their gelatinolytic. Using a caseinolytic assay. Both proteins were shown to be completely inhibited with metal chelators such as EDTA, 1,10-phenanthroline and citrate.

The gelatinolytic activity was analysed by zymography. On the gelatin zymogram (0.25% gelatin), RVV-X showed a clear band but Daborhagin did not show any gelatin degradation (Figure 6.11).

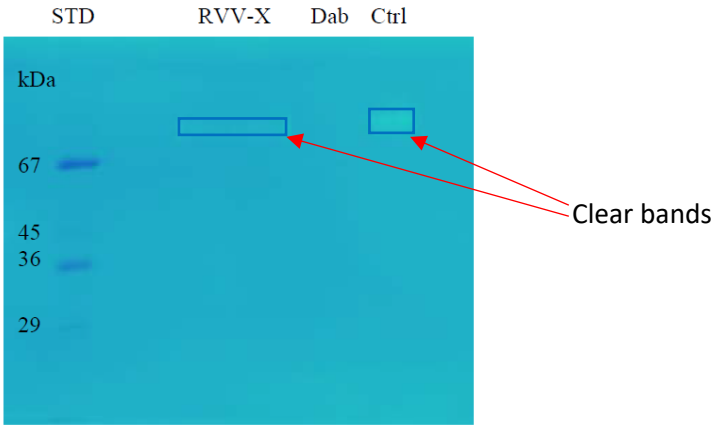


Figure 6.11. Gelatinolytic activity of RVV-X and daborhagin on 0.25% gelatin zymogram. The gelatinolytic activity of the enzyme is defined as a clear band on the SDS-polyacrylamide gel co-polymerised with gelatin after 48 hour-incubation at 37°C. Dab: Daborhagin; Ctrl: combined sample of two purified proteins. RVV-X, but not Daborhagin, showed gelatinolytic activity.

6.3.3.5. Fibrinogenolytic activity

The fibrinogenolytic activity of the two proteins was determined using 12% SDS-PAGE after incubation with fibrinogen solution for different times at 37°C. Daborhagin digested the α-chain of human fibrinogen within 1 hour of incubation. RVV-X only revealed fibrinogenolytic activity after an overnight incubation (Figure 6.12).

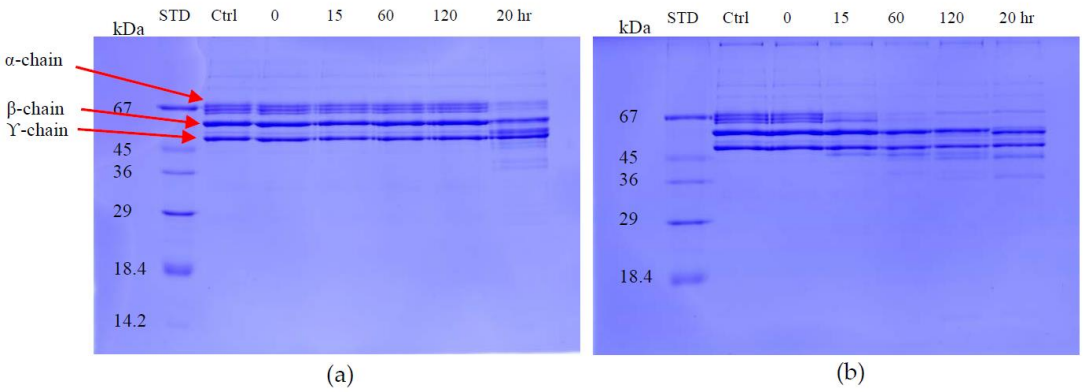


Figure 6.12. Fibrinogenolytic activity of RVV-X and daborhagin. 10 µg/mL purified enzyme was incubated with 1 mg/mL fibrinogen solution at 0, 15, 60, 120 minutes and 20 hour-incubation. Sample: (a) RVV-X; (b) daborhagin. Ctrl: fibrinogen control. STD: molecular weight markers.

6.3.4. Inhibition of SVMPs with synthetic tripeptides

6.3.4.1. Effect of synthetic tripeptides on the gelatinolytic activity of RVV-X

The gelatinolytic activity of RVV-X was completely inhibited by both synthetic tripeptides pEKW and pERW at 5 mM concentration when incubated with 1 mg/mL gelatin solution at 37°C (Figure 6.13). The α-chains (100 kDa & 130 kDa), β-chain (200 kDa) and γ-chain (300 kDa) of gelatin were totally degraded by RVV-X in a 20 hour-incubation, whereas these gelatin subunits were still intact in samples containing tripeptides or EDTA after 20 hours of incubation. The tripeptide pEEW was included in the assay to test the specificity of amino acid residue in the second position of the tripeptides.

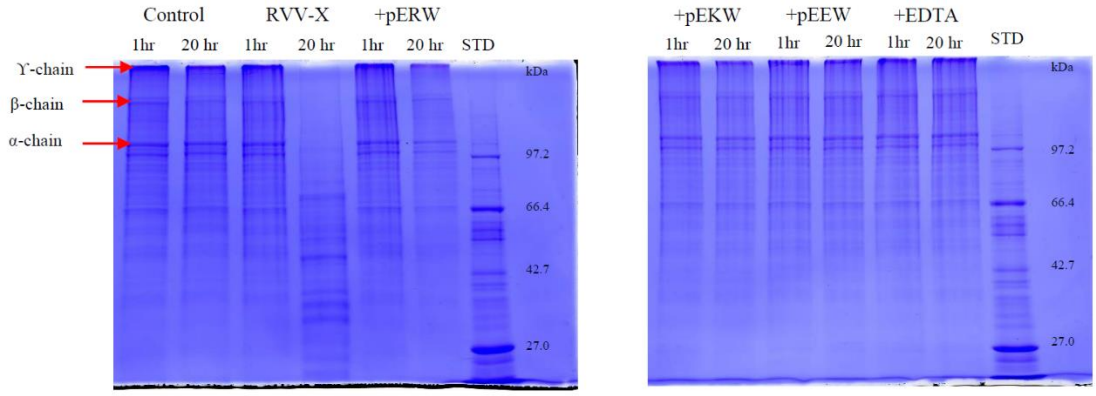


Figure 6.13. Effect of synthetic tripeptides on gelatinolytic activity of RVV-X. Gelatin (1 mg/mL) was incubated with 10 µg/mL RVV-X for 1 hour and 20 hours at 37°C, either with or without EDTA or synthetic tripeptides: pERW, pEKW, pEEW. Control = reduced gelatin, STD = molecular weight markers.

6.3.4.2. Effect of synthetic tripeptides on the fibrinogenolytic activity of daborhagin

The fibrinogenolytic activity of daborhagin was completely inhibited by both synthetic tripeptides pEKW and pERW at 5 mM concentration when incubated with 1 mg/mL fibrinogen at 37°C (Figure 6.14).

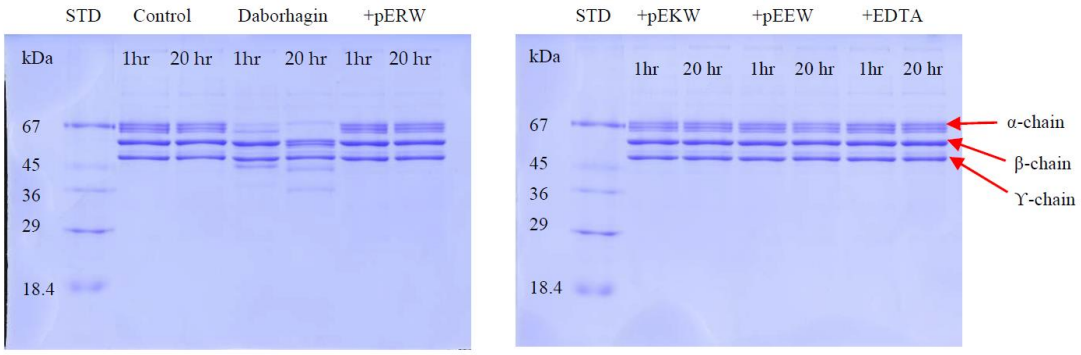


Figure 6.14. Effect of synthetic tripeptides on fibrinogenolytic activity of daborhagin. Fibrinogen (1 mg/mL) was incubated with 10 µg/mL daborhagin for 1 hour and 20 hours at 37°C with or without EDTA or synthetic tripeptides: pERW, pEKW, pEEW. Control = reduced fibrinogen, STD = molecular weight markers.

The synthetic tripeptides pERW, pEKW and pEEW showed complete inhibition of the gelatinolytic activity of RVV-X and of the fibrinogenolytic activity of daborhagin at 5

mM concentration of each inhibitor. Non-selective inhibition of all three synthetic peptides on biological activities of SVMPs reflects the importance of the first pyroGlu and the final tryptophan residue in the blocking mechanism at the active site of SVMP. The crystal structure of TM-3 (a SVMP from *Trimeresurus mucrosquamatus*) bound to tripeptide inhibitors (a proteinase and inhibitors model from Taiwan habu) revealed that the inhibitor Trp residue deeply inserts into the S-1 pocket of the protease and provides a greater inhibition than other smaller amino acids. Similarly, the pyro-ring of the inhibitor is required for fitting into the S-3 position of the protease and the activity of inhibitor becomes weaker in the absence of pyro-ring. The native middle residue is also position-specific to the S-2 site (108). Tripeptides from different species share same first (pyroGlu-) and third (tryptophan) residues. The variability of the middle residue may be dependent on species variation of the SVMPs.

6.4. Conclusions

Two major SVMPs and two endogenous tripeptides were isolated and identified from Myanmar Russell's viper venom. The two synthetic tripeptides showed specific inhibition against the fibrinogenolytic and gelatinolytic activities of the SVMPs. These findings may provide means to explore potential drug design in using these tripeptide inhibitors, or analogues of theses as alternative or additional tools in treating the toxic effects of envenomation, as well as in thrombosis.

The methods and approach used in this chapter, namely the chromatographic study on venom, characterized the properties of the isolated toxin protein in its natural, functional form. Therefore, the overall activity and composition of this toxin group, SVMPs, in venom were calculated and the biologically-active proteins isolated by this approach is the most appropriate to test their biological functions *in vitro* or *in vivo*. The disadvantage of this approach is that a large amount of venom and considerable time are needed to characterize each individual toxin. The overall identification and quantification of all toxins in a snapshot, which cannot be achieved by chromatographic studies alone, is achieved by 2D SDS-PAGE, and mass spectrometry in the proteomic study.

Chapter 7. General discussion, key conclusions and future perspectives

7.1. Summary of approaches

The aim of this thesis was to investigate the snake venom metalloproteinases and their inhibitory tripeptides from Myanmar Russell's viper with respect to two specific approaches: the analysis of their transcripts in a venom transcriptome, and purification and characterisation of them from the venom. For the transcriptomic study, a cDNA library of the male venom gland was constructed using the CloneMinerTM II cDNA Library Construction Kit. The cDNA insert lengths obtained from the library were in the range 107 to 960 bp and included a few clones corresponding to known venom toxins. The clone for the snake venom metalloproteinase covered a piece of sequence of approximately 400 bp length. The 5'RACE products of that metalloproteinase clone have been identified as a new distintegrin belonging to Myanmar Russell's viper. Thus, an exact annotation of the cloned gene by BLAST comparison was not possible. An increase in coverage of a longer sequence will be required for its more precise annotation.

High-throughput next-generation sequencing when used in conjunction with *de novo* assembly, was proven to be successful for toxin gene analysis in both quantitative and qualitative ways. Thus, RNA-Seq of two libraries (one contained 2 male snakes and the other contained 1 female snake tissue) was performed on Illumina HiSeq2000 platform. The average read length for the male library was 281 bp and for the female was 284 bp in size. The total size of reads from the male was 0.82 Gb, and from the female it was 0.79 Gb. Since Myanmar Russell's is a non-model organism, Trinity *de novo* assembly software was used. The largest contigs assembled by Trinity were 31000 bp and 17000 bp for male and female libraries respectively. Therefore, NGS sequencing with *de novo* assembly supported a more comprehensive coverage of Myanmar Russell's viper transcriptomes.

Incredibly large datasets from NGS require the right bioinformatic tools to achieve the comprehensive toxin cataloging to search for individual toxins or isoforms within one toxin family. Functional annotation of genes from non-model organisms is based mainly on homology to annotated genes of model organisms. It is a particularly difficult task in the case of snake venom as many toxin genes were recruited from

genes of non-venomous functions and do not have proper functional orthologues in other species. We have used manual annotation of nucleotide sequences retrieved from NCBI database by search with 'Serpents' and 'venom' as key words. The toxin sequences were grouped according to the annotated toxin names. Then, metalloprotease contigs were extracted from the assembled transcriptome by application of Phython software. The SVMPs contigs were later clustered by their gi numbers. Further analysis with Blastx search of these contigs finally helped to identify the different toxin gene isoforms presented in the SVMP family. The presence of a signal peptide and signature motifs unique to the SVMP family were useful for manual cataloguing of the transcripts. Further sequence analysis to identify gene coding regions (Coding DNA Sequence) and a check for nucleotide substitutions, deletions, and/or additions in the coding regions need to be completed and this is in our future research plans.

In addition to documenting the SVMP expression profile from Myanmar Russell's viper venom gland, the novel new SVMPI transcripts were identified. From these, we predicted the amino acid residues of tripeptides which are likely to be secreted in the venom.

In second part of this research, snake venom metalloproteinases and endogenous tripeptides were purified using a 3 step chromatography procedure for the major SVMPs and 2 step chromatography for tripeptides. The identification of purified toxin proteins and tripeptides required higher sensitivity and higher resolution mass spectrometry. Firstly, the purified proteins were digested by trypsin and then ionized by soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Consequently, to determine the order of the amino acid residues, fragmented peptides were further broken up to yield a spectrum of daughter ions to detect their m/z (mass/charge) values using tandem mass spectrometry (MS/MS). The mass spectrum was then matched against the database to identify the toxins and tripeptides. Thus, analysis of peptides and protein components of venoms undoubtedly promotes structure-function studies of individual toxins as well as for a discovery of novel biomolecules for therapeutic use.

7.2. Key conclusions

- From conventional cDNA library of male Myanmar Russell's viper venom gland, two full-length toxin sequences: phospholipase A_2 (PLAs) and cysteine-rich secretory protein (CRISP) were identified.
- A disintegrin sequence was obtained from 5'RACE reaction using primers designed on basis of partial sequence information that have annotation of metalloproteinase from above cDNA library.
- RNA-Sequencing of Myanmar Russell's viper venom glands generated 0.82 Gb and 0.79 Gb of reads in male and female samples respectively. *De novo* assembly software Trinity performed better assembly than CLC Genomics WorkBench.
- The highly-expressed toxins in Myanmar Russell's viper transcriptome are metalloproteinases, C-type lectin like protiens, Phospholipase A_2, serine proteases and vascular endothelial growth factor (VEGF).
- The expression profile of SVMP transcripts from MRV demonstrates potential differences in toxicity of male- and female-snake venoms. The relative abundance of SVMP transcripts is similar between male and female venom gland transcriptomes. Disintegrin transcripts are highly expressed followed by P-III SVMP and P-II SVMP transcripts.
- Cloning of a sequence of the RVV-X heavy chain unexpectedly lead to a discovery of a new truncated form of P-III SVMP and suggested a new way of evolution of the snake venom composition.
- PCR products of both RVV-X heavy chain and disintegrin produced with same primer-pairs method suggested that these two genes are likely to come from the same ancestor unless it is proved otherwise by their genomic sequences.
- Sequence alignment of RVV-X heavy chains and light chains from different sub-species of Russell's vipers showed that the variation in clinical manifestations concerning the blood coagulation of snake sub-species might depend on a structure/function of the RVV-X light chains.
- Results of the analysis of new SVMPI transcripts from MRV transcriptome suggests an existence of inhibitor tripeptides in crude venom.

- In Myanmar Russell's viper, SVMPs contribute to 20% of the crude venom and P-III SVMPs were found to be the major venom component.
- Two major P-III SVMPs, RVV-X and daborhagin were purified. RVV-X has a specific proteolytic activity on gelatin and daborhagin is a potent α-fibrinogenase.
- Two tripeptide inhibitors, pEKW and pERW were purified and identified from MRV. From which, the tripeptide pERW is a novel one.
- The synthetic tripeptides, pEKW and pERW were shown to completely inhibit the gelatinolytic and fibrinogenolytic activity of respective SVMPs at 5 mM concentration.

Haemotoxicity is a predominant symptom caused by Russell's vipers. The haemotoxic manifestations in humans, such as circulatory shock, spontaneous bleeding, and acute renal failure, lead to death following viper bite envenoming. The induction of shock is a multifactorial event, and it is exacerbated by internal haemorrhage. Spontaneous bleeding is mainly caused by direct damage to capillary endothelium, as well as due to defibrination. Severe fibrinogenemia, formation of fibrin degradation products (FDP) are caused by venom-activation of the blood coagulation cascade. Swelling and oedema formation are consequences of venom-induced increased vascular permeability. Thus, acute renal failure may occur due to hypotension causing reduced renal blood flow and fibrin deposition in glomerular capillaries (165).

In a healthy person, the hemostatic system is based on complex interactions between plasma proteins (coagulation factors), endothelial cells and blood platelets (Figure 7.1). Snake venom components acts on these components to disturb the normal hemostasis mechanisms, resulting in intravascular coagulation, increased vascular permeability and impaired platelet aggregation.

Spontaneous bleeding into vital organs is mainly caused by direct damage to capillary endothelium by haemorrhagic venom components, such as SVMPs, SVSPs and disintegrins. Phospholipases A_2 may increase vascular permeability indirectly the release of inflammatory mediators, such as arachidonic acid, prostaglandins, leukotrienes and histamines.

Procoagulants, such as RVV-X (SMP), RVV-V (SVSP) generate thrombin by activating prothrombin, causing severe disseminated intravascular coagulation (DIC), which is a contributing factor to death in viper bite patients. Some SVSPs play a role in anticoagulation by activating protein C, which in turn inactivates Factor Va and VIIIa to slow down blood coagulation.

SVMPs from snake venom also disturb the hemostatic system by exhibiting α- or β-fibrinogenase activities or by inhibiting platelet aggregation. SVSPs act as fibrinogenases, as well as fibrinases or potent platelet aggregation molecules. C-type lectins and disintegrins modulate platelet aggregation to impair platelet function. Phospholipases A_2 activate and/or inhibit platelet aggregation by hydrolysis of, or by directly binding to platelet phospholipids and consequently has an anticoagulant effect. Thrombocytopenia and poor clot retraction caused by impaired platelet aggregation, can also potentiate the bleeding (166-169). .

The geographical, age-dependent, and even individual difference in venom composition and isoforms of each toxin may influence the outcome of snake-bite envenomations. From results of this thesis, SVMPs are the most highly-expressed, with 15 isoforms in Myanmar Russell's viper glands, reflecting that this is the major component of the venom. Disintegrins, RVV-X and DSAIP, which are also highly-expressed, are key contributors for coagulopathy, severe haemorrhage, renal failure, oedema, shock and pituitary infarction. Purification of MRV via column chromatography showed that SVMPs comprise a higher percentage (20%) in this venom than in venom of Russell's viper from other countries, such as Southern India (9.5%) and Sri Lanka (6.9%). In this thesis, it was not possible to compare the results with components of Thailand Russell's viper, as the composition of SVMPs in Thailand RV venom is not defined. However, in a cDNA library of Thai RV venom gland, the most abundant toxin transcripts were phospholipase A_2 (63). Suwansrinon, et al (2007) study (27) showed that Thailand Russell's viper exhibited high phospholipase A_2 activity and low proteolytic activity. Moreover, the renal haemodynamic changes in dogs induced by Thai Russell's viper venom appeared to correlate better with proteolytic enzyme activity than PLA_2 activity. Thus, in Thai Russell's viper species, RVV-X contributes to coagulopathy and renal failure, but this is not very severe

enough to lead to pituitary infarction. Signs and symptoms, such as oedema and shock, are unlikely to occur in Russell's viper bite in Thailand, and this might be because of low level or absence of SVMPs/Disintegrins, which directly act on endothelial cells and platelet aggregation. In addition, phospholipase A_2 has an indirect effect on endothelial cells and platelets, that could explain the less-severe clinical outcomes than in Myanmar Russell's viper bite. The significant haemolytic feature in Thai Russell's viper bite might be due to direct hemolysis effects of phospholipase A_2 on red blood cells' phospholipids membrane.

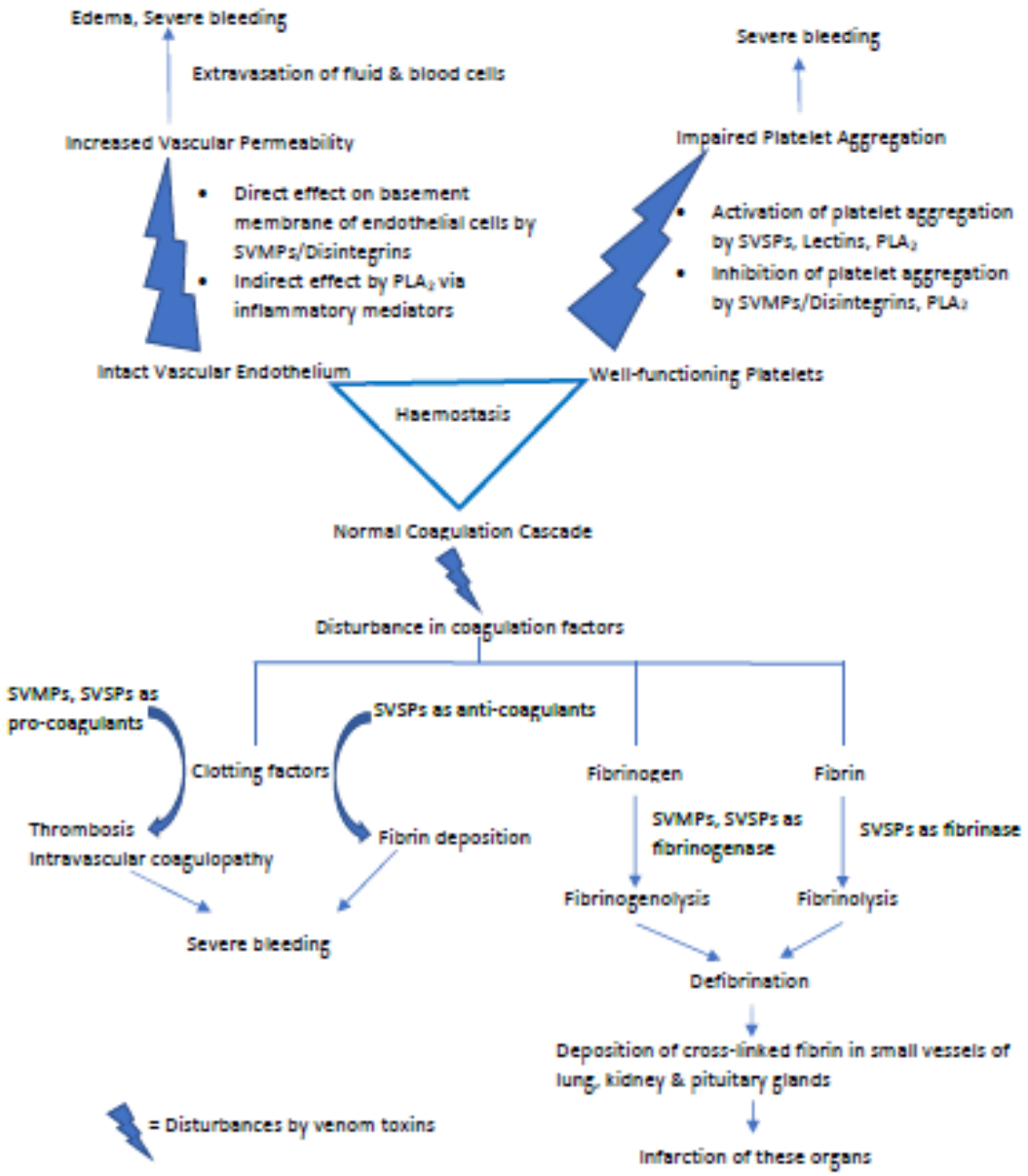


Figure 7.1. Effects of snake toxins on hemostatic system and their consequences in human. SVMPs = snake venom metalloproteinases, SVSPs = snake venom serine proteinases, PLA_2 = Phospholipase A_2. Based on Meier & Stocker, 1991; Hutton & Warrell, 1993; Rojnuckarin, 2008; Kini, 2011; Fatima, & Fatah, 2014; and Yee, 2017.

7.3. Future perspectives

The overall aims of this thesis, the identification, quantification, characterisation of SVMP transcripts from venom glands and venom, have been achieved. The high expression of SVMPs compared to other toxins, their diverse isoforms match the significant clinical outcomes in Russell's viper bite in Myanmar, compared to other countries. Further transcriptomic analysis of other toxin and non-toxin transcripts could help in detailed understanding the pathophysiology in Myanmar Russell's viper bite, the evolution of toxins as well as the new discovery of pharmacological molecules from Myanmar Russell's viper venom.

The transcriptomic analysis elucidated a sex-specific difference in venom toxicity. This finding suggests that venom used for antivenom production should include combination of venoms from both snake sexes in addition to different ages and different geographical regions of donor snakes to ensure a high coverage of anti-venom effectiveness.

The results from the assay of inhibitory activity of tripeptides against biological activities of the major SVMPs encourages the future development of a therapeutic agent for Russell's viper-bite treatment and may also be utilized in new approaches for prophylactics and treatment of thrombosis-related diseases.

References

1. Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, et al. The global burden of snakebite: A literature analysis and modelling based on regional estimates of envenoming and deaths. PLoS Med. 2008;5:e218.

2. Theakston RDG, Warrell DA. Crisis in snake antivenom supply for Africa. The Lancet. 2000;356:2104.

3. Belt PJ, Malhotra A, Thorpe RS, Warrell DA & Wuster W. Russell's viper in Indonesia: snakebite and systematics. Symposia of the Zoological Society of London. 1997;70:219-34.

4. Warrell D, Gutierrez J, Calvete JJ, Williams D. New approches & technologies of venomics to meet the challenge of human envenoming by snakebites in India. Indian Journal of Medical Research. 2013;138:38-59.

5. Warrell DA. Royal Society of Tropical Medicine and Hygiene Symposium at Manson House , London, 18 May 1989. Snake venoms in science and clinical medicine * 1 . Russell ' s viper : biology , venom and treatment of bites. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1989;83:732–40.

6. Min-Swe. Epidermiology of snakebite in Burma with special reference to Gyobingauk Township. . Yangon: Institute of Medcine I; 1977.

7. Aye-Aye-Myint, Tun-Pe, Tin-Zar-Maw. An epidemiological study of snakebite and venomous snake survey in Myanmar. In: Management of snakebite and research. Report and working papers of a seminar, Yangon, Myanmar, 11-12 December 2001. New Delhi: WHO, Regional Office for South-East Asia; 2002. p. 12-6.

8. Myint-Lwin, Warrell DA, Phillips RE, Tin-Nu-Swe, Tun-Pe, Maung-Maung-Lay. Bites by Russell's viper (Vipera russelli siamensis) in Burma: Haemostatic, vascular, and renal disturbances and response to treatment. Lancet. 1985;8467:1259-64.

9. Hung D-Z, Wu M-L., Deng J-F & Lin-Shiau S-Y. Russell's viper snakebite in Taiwan: differences from other Asian countries. Toxicon. 2002;40:1291-8.

10. Antonypillai CN, Wass JA, Warrell DA, Rajaratnam HN. Hypopituitarism following envenoming by Russell's vipers (*Daboia siamensis* and *D. russelii*) resembling Sheehan's syndrome: first case report from Sri Lanka, a review of the literature and recommendations for endocrine management. QJM. 2011;104:97-108.

11. Kularatne SAM. Epidemiology and clinical picture of the russell's viper (*Daboia russelii russelii*) bite in Anuradhapura, Sri Lanka: A prospective study of 336 patients. The Southeast Asian Journal of Tropical Medicine and Public Health. 2003;34:855-62.

12. Kularatne SA, Silva A, Weerakoon K, Maduwage K, Walathara C, Paranagama R, et al. Revisiting Russell's viper (*Daboia russelii*) bite in Sri Lanka: Is abdominal pain an early feature of systemic envenoming? PLoS One. 2014;9:e90198.

13. Warrell DA. Snake venoms in science and clinical medicine 1. Russell's viper: biology, venom and treatment of bites. Transactions of The Royal Society of Tropical Medicine and Hygiene. 1989;83:732-40.

14. Leon G, Sanchez L, Hernandez A, Villalta M, Herrera M, Segura A, Estrada R & Gutierrez JM. Immune response towards snake venoms. Inflammation & Allergy - Drug Targets. 2011;10:1-18.

15. Russell FE. Snake Venom Posioning. Venoms. Philadelphia: Lippincott; 1980. p. 139-234.

16. Tu AT. Overview of Snake Venom Chemistry. In: Singh BR, Tu AT, editors. Natural Toxin II. New York: Plenum Press; 1996. p. 37-62.

17. Risch M, Georgieva D, von Bergen M, Jehmlich N, Genov N, Arni RK, et al. Snake venomics of the Siamese Russell's viper (*Daboia russelli siamensis*) -- relation to pharmacological activities. Journal of Proteomics. 2009;72:256-69.

18. Sharma M, Das D, Iyer JK, Kini RM, Doley R. Unveiling the complexities of Daboia russelii venom, a medically important snake of India, by tandem mass spectrometry. Toxicon. 2015;107:266-81.

19. Kalita B, Patra A, Mukherjee AK. Unraveling the proteome composition and immuno-profiling of Western India Russell's viper venom for In-depth understanding of Its pharmacological properties, clinical manifestations, and effective antivenom treatment. Journal of Proteome Research. 2017;16:583-98.

20. Tan NH, Fung SY, Tan KY, Yap MK, Gnanathasan CA, Tan CH. Functional venomics of the Sri Lankan Russell's viper (*Daboia russelii*) and its toxinological correlations. Journal of Proteomics. 2015;128:403-23.

21. Nawarak J, Sinchaikul S, Wu CY, Liau MY, Phutrakul S, Chen ST. Proteomics of snake venoms from Elapidae and Viperidae families by multidimensional chromatographic methods. Electrophoresis. 2003;24:2838-54.

22. Calvete JJ, Escalante T, Sanz L. Snake venomics of Bitis species reveals large intragenus venom composition variation: application to taxonomy of congeneric taxa. Journal of Proteome Research. 2007;6:2732-45.

23. Georgieva D, Risch M, Kardas A, Buck F, von Bergen M, Betzel C. Comparative analysis of the venom proteomes of *Vipera ammodytes ammodytes* and *Vipera ammodytes meridionalis*. Journal of Proteome Research. 2008;7:866-86.

24. Prasad NB, Uma B, Bhatt SKG, Gowda VT. Comparative characterisation of Russell's viper (*Daboia/Vipera russelli*) venoms from different regions of the Indian peninsula. Biochimica et Biophysica Acta. 1999;1428:121-36.

25. Casewell NR, Wagstaff SC, Wuster W, Cook DA, Bolton FM, King SI, et al. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proceedings of the National Academy of Sciences of the United States of America. 2014;111:9205-10.

26. Khin-Than-Yee, Khow O, Noiphrom J, Aung-Myat-Kyaw, Lwin-Zar-Maw, May-Thu-Kyaw & Chulasugandha P. Purification and characterization of metalloproteinase from Myanmar Russell's viper (*Vipera russelii*) venom. The Myanmar Health Sciences Research Journal. 2014;26:93-102.

27. Suwansrinon K, Khow O, Mitmoonpitak C, Daviratanasilpa S, Chaiyabutr N, Sitprija V. Effects of Russell's viper venom fractions on systemic and renal hemodynamics. Toxicon. 2007;49:82-8.

28. Lu Q, Clemetson JM & Clemetson KJ. Snake venoms and hemostasis. Journal of Thrombosis and Haemostasis. 2005;3:1791-9.

29. Takeya H, Nishida S, Miyata T, Kawada S-I, Saisaka Y, Morita T, et al. Coagulation factor X activating enzyme from Russell's viper venom (RVV-X). The Journal of Biological Chemistry. 1992;267:14109-17.

30. Suntravat M, Yusuksawad M, Sereemaspun A, Perez JC, Nuchprayoon I. Effect of purified Russell's viper venom-factor X activator (RVV-X) on renal hemodynamics, renal functions, and coagulopathy in rats. Toxicon. 2011;58:230-8.

31. Chakrabarty D, Datta K, Gomes A, Bhattacharyya D. Haemorrhagic protein of Russell's Viper venom with fibrinolytic and esterolytic activities. Toxicon. 2000;38:1475-90.

32. Chakrabarty D, Bhattacharyya D, Sarkar H, Lahiri S. Purification and partial characterization of a haemorrhagin (VRH-I) from *Vipera russelli russelli* venom. Toxicon. 1993;31:1601-14.

33. Mukherjee AK. Characterization of a novel pro-coagulant metalloprotease (RVBCMP) possessing alpha-fibrinogenase and tissue haemorrhagic activity from venom of *Daboia russelli russelli* (Russell's viper): evidence of distinct coagulant and haemorrhagic sites in RVBCMP. Toxicon. 2008;51:923-33.

34. Chen HS, Tsai HY, Wang YM, Tsai IH. P-III hemorrhagic metalloproteinases from Russell's viper venom: cloning, characterization, phylogenetic and functional site analyses. Biochimie. 2008;90:1486-98.

35. Tokunaga F, Nagasawa K, Tamura S, Miyata T, Iwanaga S & Kisie W. The factor V-activating enzyme (RVV-V) from Russell's viper venom. The Journal of Biological Chemistry. 1988;263:17471-81.

36. Mukherjee AK, Mackessy SP. Biochemical and pharmacological properties of a new thrombin-like serine protease (Russelobin) from the venom of Russell's Viper (*Daboia russelii russelii*) and assessment of its therapeutic potential. Biochimica et Biophysica Acta. 2013;1830:3476-88.

37. Mukherjee AK. The pro-coagulant fibrinogenolytic serine protease isoenzymes purified from *Daboia russelii russelii* venom coagulate the blood through factor V activation: role of glycosylation on enzymatic activity. PLoS One. 2014;9:e86823.

38. Maung-Maung-Thwin, Gopalakrishnakone P, Yuen R & Tan CH. A major lethal factor of the venom of Burmese Russell's viper (*Daboia russelli siamensis*): Isolation, N-terminal sequencing and biological activities of daboiatoxin. Toxicon. 1995;33:63-76.

39. Prasad NB, Kemparaju K, Bhatt KGS & Gowda TV. A platelet aggregation inhibitor phospholipase A_2 from Russell's viper (*Vipera russelli*) venom: Isolation and characterization. Toxicon. 1996;34:1173-85.

40. Jayanthi GP, Kasturi, S. & Gowda, T. V. Dissociation of catalytic activity and neurotoxicity of a basic phospholipase A_2 from Russell's viper (*Vipera russelli*) venom. Toxicon. 1989;27:875-85.

41. Vishwanath BS, Kini RM & Gowda TV. Purification and partial biochemical characterization of an edema inducing phospholipase A_2 from *Vipera russelli* (Russell's viper) snake venom. Toxicon. 1988;26:713-20.

42. Kasturi S, Gowda TV. Purification and characterization of a major Phospholipase A_2 from Russell's viper (*Vipera russelli*) venom. Toxicon. 1989;27:229-37.

43. Wang Y-M, Lu P-J, Ho C-L, Tsai I-H. Characterization and molecular cloning od neurotoxic phospholipase A_2 from Taiwan viper (*Vipera russelli formosensis*). European Journal of Biochemistry. 1992;209:635-41.

44. Maity G, Mandal S, Chatterjee A, Bhattacharyya D. Purification and characterization of a low molecular weight multifunctional cytotoxic phospholipase A_2 from Russell's viper venom. Journal of Chromatography B 2007;845:232-43.

45. Sharma M, Iyer JK, Shih N, Majumder M, Mattaparthi VS, Mukhopadhyay R, et al. Daboxin P, a major phospholipase A_2 enzyme from the Indian *Daboia russelii russelii* venom targets Factor X and Factor Xa for Its anticoagulant activity. PLoS One. 2016;11:e0153770.

46. Mukherjee AK, Dutta S, Kalita B, Jha DK, Deb P, Mackessy SP. Structural and functional characterization of complex formation between two Kunitz-type serine protease inhibitors from Russell's Viper venom. Biochimie. 2016;128-129:138-47.

47. Cheng AC, Tsai IH. Functional characterization of a slow and tight-binding inhibitor of plasmin isolated from Russell's viper venom. Biochimica et Biophysica Acta. 2014;1840:153-9.

48. Warrell DA. Snake bite. Lancet. 2010;375:77-88.

49. Markland FS. Snake venoms and the hemostatic system. Toxicon. 1998;36:1749-800.

50. Koyama J-i, Inoue S, Ikeda K & Hayashi K. Purification and amino-acid sequence of a nerve growth factor from the venom of *Vipera russelli russelli*. Biochimica et Biophysica Acta. 1992;1160:287-92.

51. Mukherjee AK, Dutta S, Mackessy SP. A new C-type lectin (RVsnaclec) purified from venom of *Daboia russelii russelii* shows anticoagulant activity via inhibition of FXa and concentration-dependent differential response to platelets in a Ca^{2+}-independent manner. Thrombosis Research. 2014;134:1150-6.

52. Pathan J, Mondal S, Sarkar A, Chakrabarty D. Daboialectin, a C-type lectin from Russell's viper venom induces cytoskeletal damage and apoptosis in human lung cancer cells in vitro. Toxicon. 2017;127:11-21.

53. Yamazaki Y, Takani K, Atoda H, Morita T. Snake venom vascular endothelial growth factors (VEGFs) exhibit potent activity through their specific recognition of KDR (VEGF receptor 2). The Journal of Biological Chemistry. 2003;278:51985-8.

54. Tin-Win, Aye-Kyaw, Sanda, San-Aye, Win-Aung, Khin-Pa-Pa-Kyaw & Aung-Myat-Kyaw. Purification, characterization and biological properties of phosphodiesterase from Russell's viper (*Vipera russelli*) venom. The Snake. 1998;28:83-9.

55. Mitra J, Bhattacharyya D. Phosphodiesterase from *Daboia russelli russelli* venom: purification, partial characterization and inhibition of platelet aggregation. Toxicon. 2014;88:1-10.

56. Hla-Kathy-Win. Studies on phosphomonodiesterase of Russell's viper venom (*Vipera russelli siamensis*). Yangon, Myanmar: University of Yangon; 1996.

57. Chen HS, Wang YM, Huang WT, Huang KF, Tsai IH. Cloning, characterization and mutagenesis of Russell's viper venom L-amino acid oxidase: Insights into its catalytic mechanism. Biochimie. 2012;94:335-44.

58. Bernheimer AW, Linder R, Weinstein SA & Kim KS. Isolation and characterization of a phospholipase B from venom of Collett's snake, *Pseudechis colletti*. Toxicon. 1987;25:547-54.

59. Brahma RK, McCleary RJ, Kini RM, Doley R. Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes. Toxicon. 2015;93:1-10.

60. Vonk FJ, Jackson K, Doley R, Madaras F, Mirtschin PJ, Vidal N. Snake venom: From fieldwork to the clinic: Recent insights into snake biology, together with new technology allowing high-throughput screening of venom, bring new hope for drug discovery. Bioessays. 2011;33:269-79.

61. Sunagar K, Morgenstern D, Reitzel AM, Moran Y. Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. Journal of Proteomics. 2016;135:62-72.

62. Junqueira-de-Azevedo IdLM, Ho PL. A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). Gene. 2002;299:279-91.

63. Sai-Ngam A. Gene Expression Analysis of Russell's Viper Venom and Molecular Cloning, Expression and Functional Study of RV Factor X Activator. Thailand: Chulalongkorn University; 2007.

64. Rokyta D, Lemmon A, Margres M, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics. 2012;13:312.

65. Hughes A. The evolution of functionally novel proteins after gene duplication. Proceedings: Biological Sciences. 1994;256:119-24.

66. Wong ES, Belov K. Venom evolution through gene duplications. Gene. 2012;496:1-7.

67. Kordis D. Evolution of phospholipase A_2 toxins in venomous animals. Acta Chimica Slovenica. 2011;58:638-46.

68. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wuster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Molecular Biology and Evolution. 2011;28:2637-49.

69. Sanz L, Harrison RA, Calvete JJ. First draft of the genomic organization of a PIII-SVMP gene. Toxicon. 2012;60:455-69.

70. Juarez P, Comas I, Gonzalez-Candelas F, Calvete JJ. Evolution of snake venom disintegrins by positive Darwinian selection. Molecular Biology and Evolution. 2008;25:2391-407.

71. Gomis-Ruth F, Kress L, Bode W. First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenes. The EMBO Journal. 1993;12:4151-7.

72. Bjarnason J, Fox J. Snake venom metalloendopeptidases: reprolysins. Methods in Enzymology. 1995;248:345-68.

73. Casewell NR. On the ancestral recruitment of metalloproteinases into the venom of snakes. Toxicon. 2012;60:449-54.

74. Takeda S. ADAM and ADAMTS family proteins and snake venom metalloproteinases: A structural overview. Toxins 2016;8.

75. Seals DF, Courtneidge S. The ADAMs family of metalloproteinases: multidomain proteins with multiple functions. Genes and Development. 2003;17:7-30.

76. Calvete JJ, Juarez P, Sanz L. Snake venomics. Strategy and applications. Journal of Mass Spectrometry. 2007;42:1405-14.

77. Markland FS, Swenson S. Snake venom metalloproteinases. Toxicon. 2013;62:3-18.

78. Fox JW, Serrano SM. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. FEBS J. 2008;275:3016-30.

79. Fox JW, Serrano SM. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon. 2005;45:969-85.

80. Wallnoefer H, Lingott T, Gutierrez J, Merfort I, Liedl K. Backbone flexibility controls the activity and specificity of a protein-protein interface: specificity in snake venom metalloproteinases. Journal of the American Chemical Society. 2010;132:10330-7.

81. Paine M, Desmond H, Theakston R, Crampton J. Pirification, cloning, and molecular characterization of a high molecular weight haemorrhagic metalloproteinase, jararhagin, from *Bothrops jararaca* venom. The Journal of Biological Chemistry. 1992;267:22869-76.

82. Selistre-de-Araujo H, Ownby C. Molecular cloning and sequence analysis of cDNAs from metalloproteinase from broad-banded copperhead *Agkistrodon contortrix laticinctus*. Archives of Biochemistry and Biophysics. 1995;320:141-8.

83. Garcia LT, Parreiras e Silva LT, Ramos OH, Carmona AK, Bersanetti PA, Selistre-de-Araujo HS. The effect of post-translational modifications on the hemorrhagic activity of snake

venom metalloproteinases. Comparative Biochemistry and Physiology, Part C 2004;138:23-32.

84. Huang K, Holt J, Lukasiewicz H, Niewiarowski S. A low molecular weight peptide inhibiting fibrinogen interaction with platelet receptors expressed on glycoprotein IIb-IIIa complex. Journal of Biological Chemistry. 1987;262:16157-63.

85. Calvete JJ, Marcinkiewicz C, Monleon D, Esteve V, Celda B, Juarez P, et al. Snake venom disintegrins: evolution of structure and function. Toxicon. 2005;45:1063-74.

86. Kini RM, Evans HJ. Strucutral domains in venom proteins: evidence that metalloproteinases and non-enymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon. 1992;30:265-93.

87. Muniz JR, Ambrosio AL, Selistre-de-Araujo HS, Cominetti MR, Moura-da-Silva AM, Oliva G, et al. The three-dimensional structure of bothropasin, the main hemorrhagic factor from *Bothrops jararaca* venom: insights for a new classification of snake venom metalloprotease subgroups. Toxicon. 2008;52:807-16.

88. Takeda S, Igarashi T, Mori H, Araki S. Crystal strucutres of VAP1 reveal ADAMs' MDC domain architecture and its unique C-shaped scaffold. The EMBO Journal. 2006;25:2388-96.

89. Kamiguti A, Hay C, Zuzel M. Inhibition of collagen-induced platelet aggregation as the result of cleavage of α2β1-integrin by the snake venom metalloproteinase jararhagin. Biochemical Journal. 1996;320:635-41.

90. Tanjoni I, Evangelista K, Della-Casa MS, Butera D, Magalhaes GS, Baldo C, et al. Different regions of the class P-III snake venom metalloproteinase jararhagin are involved in binding to α2β1 integrin and collagen. Toxicon. 2010;55:1093-9.

91. Morita T. Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. Toxicon. 2005;45:1099-114.

92. Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M. Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. Toxicon. 2005;45:1-14.

93. Yamada D, Sekiya F, Morita T. Isolation and characteriztion of carinactivase, a novel prothrombin activator in *Echis carinatus* venom with a unique catalytic mechnism. Journal of Biological Chemistry. 1996;271:5200-7.

94. Mizuno H, Fujimoto Z, Atoda H, Morita T. Crystal structure of an anticoagulant protein in complex with the Gla domain of factor X. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:7230-4.

95. Takeda S, Igarashi T, Mori H. Crystal structure of RVV-X: an example of evolutionary gain of specificity by ADAM proteinases. FEBS Letters. 2007;581:5859-64.

96. Gowda DC, Jackson CM, Hensley P, Davidson EA. Factor X-activating glycoprotein of Russell's viper venom. The Journal of Biological Chemistry. 1994;269:10644-50.

97. Chen HS, Chen JM, Lin CW, Khoo KH, Tsai IH. New insights into the functions and N-glycan structures of factor X activator from Russell's viper venom. The FEBS Journal. 2008;275:3944-58.

98. Suzuki M, Itoh T, Bandaranayake BMAIK, Ranasinghe JGS, Athauda SBP & Moriyama A. Molecular diversity in venom proteins of the Russell's viper (*Daboia russellii russellii*) and the Indian cobra (*Naja naja*) in Sri Lanka. Biomedical Research. 2010;31:71-81.

99. Bastos VA, Gomes-Neto F, Perales J, Neves-Ferreira AG, Valente RH. Natural inhibitors of snake venom metalloendopeptidases: History and current challenges. Toxins 2016;8.

100. Omori-Satoh T, Sadahiro S, Ohsaka A, Murata R. Purification and characterization of an antihemorrhagic factor in the serum of *Trimeresurus flavoviridis*, a crotalid. Biochimica et Biophysica Acta. 1972;285:414-26.

101. Deshimaru M, Tanaka C, Fujino K, Aoki N, Terada S, Hattori S, et al. Properties and cDNA cloning of an antihemorrhagic factor (HSF) purified from the serum of *Trimeresurus flavoviridis*. Toxicon. 2005;46:937-45.

102. Valente RH, Dragulev B, Perales J, Fox J, Domont G. BJ46a, a snake venom metalloproteinases inhibitor. Isolation, characterization, cloning and insights into its mechanism of action. European Journal of Biochemistry. 2001;268:3042-52.

103. Aoki N, Tsutsumi K, Deshimaru M, Terada S. Properties and cDNA cloning of antihemorrhagic factors in sera of Chinese and Japanese mamushi (*Gloydius blomhoffi*). Toxicon. 2008;51:251-61.

104. Neves-Ferreira A, Perales J, Ovadia M, Moussatche H, Domont G. Inhibitory properties of the antibothropic complex from the South American opossum (*Didelphis marsupialis*) serum. Toxicon. 1997;35:846-63.

105. Omori-Satoh T, Yamakawa Y, Mebs D. The antihemorrhagic factor, erinacin, from the European hedgehog (*Erinaceus europaeus*), a metalloprotease inhibitor of large molecular size possessing ficolin/opsonin P35 lectin domains. Toxicon. 2000;38:1561-80.

106. Kato H, Iwanaga S, Suzuki T. The isolation and amino acid sequences of new pyroglutamanylpeptides from snake venoms. Experientia. 1966;22:49-50.

107. Huang K-F, Hung C-C, Wu S-H, Chiou S-H. Characterization of three endogenous peptide inhibitors for multiple metalloproteinases with fibrinogenolytic activity from the

venom of Taiwan Habu (*Trimeresurus mucrosquamatus*). Biochemical and Biophysical Research Communications. 1998;248:562-8.

108. Huang K-F, Chiou S-H, Ko T-P, Wang AH-J. Determinants of the inhibition of a Taiwan habu venom metalloproteinase by its endogenous inhibitors revealed by X-ray crystallography and synthetic inhibitor analogues. European Journal of Biochemistry. 2002;269:3047-56.

109. Francis B, Kaiser II. Inhibition of metalloproteinases in *Bothrops asper* venom by endogenous peptides. Toxicon. 1993;31:889-99.

110. Munekiyo SM, Mackessy SP. Presence of peptide inhibitors in rattlesnake venoms and their effects on endogenous metalloproteases. Toxicon. 2005;45:255-63.

111. Wagstaff SC, Favreau P, Cheneval O, Laing GD, Wilkinson MC, Miller RL, et al. Molecular characterisation of endogenous snake venom metalloproteinase inhibitors. Biochemical and Biophysical Research Communications. 2008;365:650-6.

112. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nature Protocols. 2006;1:581-5.

113. Platt A, Woodhall R, George JA. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. BioTechniques. 2007;43:58-62.

114. Li B, Dewey C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

115. Frohman M. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: Thermal RACE. Methods in Enzymology. 1993;218:340-56.

116. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227:680-5.

117. Heukeshoven J, Dernick R. Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. Electrophoresis. 1985;6:103-12.

118. Anson ML. The estimation of pepsin, trypsin, papain, and cathepsin with hemoglobin. Journal of General Physiology. 1938;22:79-89.

119. Folin O, Ciocalteu V. On Tyrosine and tryptophane determinations in proteins. The Journal of Biological Chemistry. 1927;73:627-50.

120. Toth M, Fridman R. Assessment of gelatinases (MMP-2 and MMP-9) by gelatin zymography. In: Brooks SA, Schumacher U, editors. Metastasis Research Protocols. Methods in Molecular Medicine. 57. Totowa: Humana Press; 2001. p. 163-73.

121. Ouyang C, Teng C-M. Fibrinogenolytic enzymes of *Trimeresurus macrosquamatus* venom. Biochimica et Biophysica Acta. 1976;420:298-308.

122. Wolfsberg T, Landsman D. Expressed Sequence Tags (ESTs). In: Baxevanis A, Ouellette B, editors. Bioiformatics: A Practical Guide to the Analysis of Genes and Proteins. 43. Second ed. New York, USA: John Wiley & Sons, Inc; 2001. p. 283-301.

123. Zelanis A, Andrade-Silva D, Rocha MM, Furtado MF, Serrano SM, Junqueira-de-Azevedo IL, et al. A transcriptomic view of the proteome variability of newborn and adult *Bothrops jararaca* snake venoms. PLoS Neglected Tropical Diseases. 2012;6:e1554.

124. Wagstaff SC, Harrison RA. 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. 2006;377:21-32.

125. Gopalan G, Thwin MM, Gopalakrishnakone P, Swaminathan K. Structural and pharmacological comparison of daboiatoxin from *Daboia russelli siamensis* with viperotoxin F and vipoxin from other vipers. Acta Crystallographica Section D Biological Crystallography. 2007;63:722-9.

126. Sai-Ngam A, Phongtananant A, Nuchprayoon I. Phospholipase A_2 genes and their expressions in Thai Russell's viper venom glands. Toxicon. 2008;52:395-9.

127. Wang Y-M, Lu P-J, Ho C-L, Tsai I-H. Characterization and molecular cloning of neurotoxic phospholipase A_2 from Taiwan viper (*Vipera russelli formosensis*). European Journal of Biochemistry. 1992;209:635-41.

128. Yamazaki Y, Morita T. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon. 2004;44:227-31.

129. Ramazanova AS, Starkov VG, Osipov AV, Ziganshin RH, Filkin SY, Tsetlin VI, et al. Cysteine-rich venom proteins from the snakes of Viperinae subfamily - molecular cloning and phylogenetic relationship. Toxicon. 2009;53:162-8.

130. Sanz L, Chen RQ, Perez A, Hilario R, Juarez P, Marcinkiewicz C, et al. cDNA cloning and functional expression of jerdostatin, a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the α1β1 integrin. The Journal of Biological Chemistry. 2005;280:40714-22.

131. Senger DR, Claffey KP, Benes JE, Perruzzi CA, Sergiou AP, Detmar M. Angiogenesis promoted by vascular endothelial growth factor: Regulation through α1β1 and α2β1 integrins. Proceedings of the National Academy of Sciences of the United States of America. 1997;94:13612-7.

132. Chakrabarty DC, C. Snake Venom Disintegrins. In: Gopalakrishnakone P, Inagaki H, Mukherjee A, Rahmy T, Vogel C-W, editors. Snake Venoms. Netherlands: Springer; 2015. p. 1-11.

133. Wheat CW. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica. 2010;138:433-51.

134. Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnology. 2008;26:1135-45.

135. Martin JA, Wang Z. Next-generation transcriptome assembly. Nature Reviews Genetics. 2011;12:671-82.

136. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: Key considerations in genomic analyses. Nature Reviews Genetics. 2014;15:121-32.

137. Menezes M, Furtado M, Travaglia-Cardoso S, Camargo A, Serrano S. Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. Toxicon. 2006;47:304-12.

138. Shetty S, Shine R. Sexual divergence in diets and morphology in Fijian sea snakes *Lanticauda colubrina* (Laticaudinae). Austral Ecology. 2002;27:77-84.

139. Sanz L, Calvete JJ. Insights into the evolution of a snake venom multi-gene family from the genomic organization of *Echis ocellatus* SVMP genes. Toxins 2016;8.

140. Sverdlov ED. Perpetually mobile footprints of ancient infections in human genome. FEBS Letters. 1998;428:1-6.

141. Makalowski W. Genomic scrap yard: How genomes utilize all that junk. Gene. 2000;259:61-7.

142. Alape-Giron A, Sanz L, Escolano J, Flores-Diaz M, Madrigal M, Sasa M, et al. Snake venomics of the lancehead pitviper *Bothrops asper*: Geographic, individual, and ontogenetic Variations. Journal of Proteome Research. 2008;7:3556-71.

143. Cardoso K, Silva M, Costa G, Torres T, Bem L, Vidal R, et al. A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu). BMC Genomics. 2010;11:605.

144. Zelanis A, Tashima A, Rocha M, Furtado M, Camargo A, Ho P, et al. Analysis of the ontogenetic variation in the venom proteome/peptidome of *Bothrops jararaca* reveals different strategies to deal with prey. Journal of Proteome Research. 2010;9:2278-91.

145. Andrade D, Abe A. Relationship of venom ontogeney and diet in bothrops. Herpetologica. 1999;55:200-4.

146. Risch M, Georgieva D, von Bergen M, Jehmlich N, Genov N, Arni RK, et al. Snake venomics of the Siamese Russell's viper (*Daboia russelli siamensis*) -- relation to pharmacological activities. Journal of Proteomics. 2009;72:256-69.

147. Okuda D, Koike H, Morita T. A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry. 2002;41:14248-54.

148. Sanz-Soler R, Lorente C, Company B, Sanz L, Juarez P, Perez A, et al. Recombinant expression of mutants of the Frankenstein disintegrin, RTS-ocellatusin. Evidence for the independent origin of RGD and KTS/RTS disintegrins. Toxicon. 2012;60:665-75.

149. Trummal K, Tonismagi K, Siigur E, Aaspollu A, Lopp A, Sillat T, et al. A novel metalloprotease from *Vipera lebetina* venom induces human endothelial cell apoptosis. Toxicon. 2005;46:46-61.

150. Kerkkamp HM, Kini RM, Pospelov AS, Vonk FJ, Henkel CV, Richardson MK. Snake genome sequencing: Results and future prospects. Toxins 2016;8.

151. Piskurek O, Austin CC, Okada N. Sauria SINEs: Novel short interspersed retroposable elements that are widespread in reptile genomes. Journal of Molecular Evolution. 2006;62:630-44.

152. Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE, et al. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. Genome Biology and Evolution. 2011;3:641-53.

153. Kordis D, Gubensek F. Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A_2 genes and in genomes of Viperidae snakes. European Journal of Biochemistry. 1997;246:772-9.

154. Tanaka Y, Oyama S, Hori S, Ushio K, Shioi N, Terada S, et al. Accelerated evolution of fetuin family proteins in *Protobothrops flavoviridis* (habu snake) serum and the discovery of an L1-like genomic element in the intronic sequence of a fetuin-encoding gene. Bioscience, Biotechnology, and Biochemistry. 2013;77:582-90.

155. Campos PF, Andrade-Silva D, Zelanis A, Paes Leme AF, Rocha MM, Menezes MC, et al. Trends in the evolution of snake toxins underscored by an integrative omics approach to profile the venom of the colubrid *Phalotris mertensi*. Genome Biology and Evolution. 2016;8:2266-87.

156. Bazaa A, Juarez P, Marrakchi N, Bel Lasfer Z, El Ayeb M, Harrison RA, et al. Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*. Journal of Molecular Evolution. 2007;64:261-71.

157. Fischer WH, Spiess J. Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. Proceedings of the National Academy of Sciences of the United States of America. 1987;84:3628-32.

158. Varro A. Posttranslational processing: Peptide hormones and neuropeptide transmitters. eLS. 2007.

159. Castoe TA, Jason de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, et al. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proceedings of the National Academy of Sciences of the United States of America. 2014;110:20645-50.

160. Hung DZ, Yu YJ, Hsu CL, Lin TJ. Antivenom treatment and renal dysfunction in Russell's viper snakebite in Taiwan: a case series. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2006;100:489-94.

161. Ding B, Xu Z, Qian C, Jiang F, Ding X, Ruan Y, et al. Antiplatelet aggregation and antithrombosis efficiency of peptides in the snake venom of *Deinagkistrodon acutus*: Isolation, identification, and evaluation. Evidence-Based Complementary and Alternative Medicine. 2015;2015:412841.

162. Chen H-S, Tsai H-Y, Wang Y-M, Tsai I-H. P-III hemorrhagic metalloproteinases from Russell's viper venom: cloning, characterization, phylogenetic and functional site analyses. Biochimie. 2008;90:1486-98.

163. Siigur E, Tonismagi K, Trummal K, Samel M, Vija H, Subbi J, et al. Factor X activitor from *Vipera lebetina* snake venom, molecular characterization and substrate specificity. Biochimica et Biophysica Acta. 2001;1568:90-8.

164. Mackessy S. Characterization of the major metalloproteinase isolated from the venom of the Northern Pacific rattlesnake, *Crotalus viridis oreganus*. Toxicon. 1996;34:1277-85.

165. Meier J, Stocker K. Effects of snake venoms on hemostasis. Critical Reviews in Toxicology. 1991;21:171-82.

166. Hutton RA, Warrell DA. Action of snake venom components on the haemostatic system. Blood Reviews. 1993;7:176-89.

167. Kini RM. Toxins in thrombosis and haemostasis: potential beyond imagination. Journal of Thrombosis and Haemostasis. 2011;9:195-208.

168. Fatima L-D, Fatah C. Pathophysiological and pharmacological effects of snake venom components: Molecular targets. Journal of Clinical Toxicology. 2014;4:1-9.

169. Rojnuckarin P. Snake venom and haemostasis - An overview. Thrombosis and Haemostasis. 2008;1:93-6.