CHAPTER 5

PROTEOMIC CHARACTERIZATION OF *NAJA SUMATRANA* (EQUATORIAL SPITTING COBRA) VENOM USING MULTIDIMENSIONAL CHROMATOGRAPHIC APPROACH

5.1 INTRODUCTION

Snake venoms are complex mixtures containing predominantly enzymatic or non-enzymatic proteins and polypeptides that possess various pharmacological actions (Doley and Kini, 2009). Recent advances in proteomics have made it possible to globally identify protein components in the venom, including those that exist only in minute amount (Calvete *et al.,* 2007; Calvete *et al*., 2009a; Calvete *et al*., 2009b; Gutiérrez *et al*., 2009; Jiang *et al.,* 2011; Nawarak *et al*., 2003; Risch *et al*., 2009 and Wagstaff *et al*., 2009). This detailed knowledge of the toxin composition of venoms will contribute not only to in-depth understanding on the pathogenesis of snake envenomation mechanism and development of highly effective antivenom but could also contribute to discovery of novel drugs and biomedical tools, as it has been well established that snake venoms contain novel toxins with potential therapeutic and diagnosis applications (Gomes *et al*., 2010; Kapoor, 2010; Koh *et al*., 2006; Ménez *et al*., 2006 and Pal *et al*., 2002). To date, proteomic characterization of venom proteins (often known as venomics) of a considerable number of medically important venomous snakes have been reported (Ching *et al.,* 2012; Georgieva *et al.,* 2008; Makran *et al.,* 2012; Petras *et al.,* 2011 and Rodrigues *et al.,* 2012).

Naja sumatrana (Equatorial spitting cobra) is listed as one of the category 1 medically important venomous snakes of Southeast Asia (WHO, 2010). It is indigenous to Peninsular Malaysia, Borneo and Sumatra (Wüster, 1996). It was known earlier as Malayan cobra (*Naja sputatrix*), however, recently *N. sputatrix* has been re-designated to spitting cobra in Java and known as Javan spitting cobra (Wüster, 1996). In the earlier studies, Tan *et al.* (1982a, 1982b and 1983) had reported isolation and structural characterization of the major toxins of the *N. sputatrix* venom. Molecular cloning studies using the mRNAs isolated from the

venom glands of *N. sputatrix* revealed the predicted amino acid sequences of 9 polypeptide neurotoxins, 5 polypeptide cardiotoxins and 3 phospholipases A_2 (Afifiyan *et al*., 1999; Armugam *et al*., 1997; Jeyaseelan *et al*., 1998 and Poh *et al*., 2002). Since the geographical source of the *N. sputatrix* venom or venom glands used in these studies were not specified, it cannot be ascertained whether the toxins characterized are toxins of true *N. sputatrix* or *N. sumatrana* venom. The chromatographic elution profiles of these two spitting cobras of confirmed a geographical difference in the protein compositions as demonstrated in Chapter 4. In the present work, shotgun analysis, ion exchange chromatographic fractionation and 2D-electrophoresis in conjunction with MALDI-TOF/TOF were used to investigate the proteome of *N. sumatrana* venom. Among all the Asiatic cobra, only the proteome of the non-spitting cobra, *Naja naja atra* (Chinese cobra) venom has been elucidated (Li *et al*., 2004) and a preliminary study of the proteome of the venom of Thai non-spitting cobra, *Naja kaouthia* (monocled cobra) was reported (Kulkeaw *et al*., 2007). A full characterization of the proteome of *N. sumatrana* venom will contribute to the understanding of the pathophysiology of *N. sumatrana* envenomation and antivenom treatment.

5.2 METHODS

5.2.1 In-solution digestion

The isolated and purified fractions from *N. sumatrana* venom were digested with trypsin (Trypsin Gold, Mass Spectrometry Grade). Fifteen microliters of 50 mM ammonium bicarbonate was mixed with 1.5 µl of 100 mM dithiothreitol (DTT) in a microtube. Ten microliters of protein solution was added into this mixture and 0.5 µl of UltraPure^{TM} water was added to make the final volume up to 27 µl. The sample was incubated at 95°C for 5 min and cooled to room temperature. Later, 3 µl of 100 mM iodoacetamide (IAA) was added into the mixture and incubated in dark at room temperature at 20 min. Trypsin was prepared in a concentration 0.1 µg/µl. One microliter of trypsin was added into the sample and incubated at 37°C for 3 h. An additional 1 µl of trypsin was added and the digestion was allowed overnight at 30°C. The peptides samples were prepared for MALDI-TOF/TOF as described in section 5.2.3.

5.2.2 In-gel tryptic digestion

5.2.2.1 Destaining for Coomassie blue stained gel plugs

Fifty microliters of destaining solution (50 mM ammonium bicarbonate in 50% (v/v) acetonitrile) was added into the gel plugs and shook until the gel plugs were cleared and immediately proceed to in-gel tryptic digestion.

5.2.2.2 Destaining for silver stained gel plugs

Destaining solution for silver stained gel plugs was prepared according to manufacturer protocols (Pierce® Silver Stain for Mass Spectrometry). Briefly, 74 µl of Silver Destain Reagent A and 245 µl of Silver Destain Reagent B mixed with 4 ml of UltraPure™ water which was sufficient to destain 10 gel plugs. Destaining solution (0.2 ml) was added twice into gel plugs and incubated for 15 min each. The destaining solution was discarded and the gel plugs were washed three times for 10 min each with 0.2 ml of wash solution (25 mM ammonium bicarbonate in 50% (v/v) acetonitrile). The gel plugs were then digested with sequencing grade trypsin.

5.2.2.3 Tryptic digestion

Gel plug was reduced by adding 150 µl of 10 mM dithiotreitol (DTT) in 100 mM ammonium bicarbonate for 30 min at 60°C and then it was allowed to cool to room temperature. The gel plugs were then alkylated by incubation in 150 µl of 55 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate for 20 min in dark. The gel plugs were then washed three times with 500 µl of acetonitrile in 100mM ammonium bicarbonate for 20 min. Later, the gel plugs were dehydrated by incubating in 50 µl of 100% (v/v) acetonitrile for 15 min with gentle shaking. The gel plugs were speed-vacuum dried in a Heto FD4 Lyophilizer (Cambridge Scientific) for 10-15 min at ambient temperature. The gel plugs were incubated in 25 µl of 6 ng/µl trypsin (Trypsin Gold, Mass Spectrometry Grade) in 50 mM ammonium bicarbonate, overnight at 37°C. After that, the digested peptides were extracted with 50 µl of 50% (v/v) acetonitrile with gentle shaking for 15 min, the liquid was transferred into a clean microtube. This was repeated by incubation of gel plugs in 50 µl of 100%

(v/v) acetonitrile for 15 min, the liquid was then transferred into the same clean microtube as previous. The digested peptides were completely speed-vacuum dried in a Heto FD4 Lyophilizer (Cambridge Scientific) at low speed and the samples were then desalted using ZipTip® Pipette tips (Milipore).

5.2.2.4 Desalting of digested peptides with ZipTip® Pipette tips

The digested peptides were reconstituted in 10 µl of 0.1% (v/v) formic acid. 10 µl of 50% (v/v) acetonitrile was aspirated into a ZipTip® Pipette tip and dispensed for 3 times. The ZipTip® Pipette tip was equilibrated with 10 µl of 0.1 % (v/v) formic acid for 3 times followed by aspiration of 10 µl of digested peptides. The ZipTip® Pipette tip was then washed with 10 µl of 0.1% (v/v) formic acid for 3 times. The desalted, digested peptides were then eluted out with 2 µl of 0.1% (v/v) formic acid in 50% (v/v) acetonitrile, this step was repeated for 3 times. The digested peptides were prepared for MALDI-TOF/TOF as described in section 5.2.3

5.2.3 Preparation of digested peptides for MALDI-TOF/TOF

The peptide samples were mixed with equal volume of α-cyano-4-hydroxycinnamic acid matrix (10 mg/ml in 50% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid). Seven microliter of the peptide samples were spotted on Opti-TOF™ LC/MALDI INSERT (123 x 81 mm) (Applied Biosystem®). The samples were allowed to dry completely before submitted for MALDI-TOF/TOF. The MALDI-TOF/TOF was performed using MDS SCIEX 4800 Plus MALDI TOF/TOF™ Analyzer (AB Sciex).

5.2.4 Fractionation of *N. sumatrana* venom by Resource® S Ion Exchange Chromatography

N. sumatrana venom fractions were isolated by Resource® S ion exchange chromatography as described in Chapter 4 (section 4.2.1).

5.2.5 C_{18} reverse-phase HPLC of fractions isolated from Resource® S Ion Exchange Chromatography

The Resource® S fractions collected were buffer exchanged and concentrated to 200 µl in 0.1% (v/v) trifluroacetic acid (TFA) using Vivaspin® 15R and applied to a LiChrospher® WP 300 C_{18} reverse-phase column (250 mm x 4 mm), using a Shimadzu LC-20AD HPLC system (Japan). The column was first pre-equilibrated with 0.1% (v/v) TFA, followed by elution using a linear gradient (0-100% of 0.1% (v/v) TFA in acetonitrile) for 60 min, at a flow rate of 1 ml/min. Absorbance at 280 nm was measured using the Shimadzu SPD-20A Prominence UV/Vis Detector (Japan). The purity of the protein fractions were checked for the homogeneity by SDS-PAGE as described in Chapter 3 (section 3.2.2). The fractions were collected manually and kept at -20°C before in-solution digestion as described in section 5.2.1. The digested peptides were subjected to MALDI-TOF/TOF as described in section 5.2.3.

5.2.6 Assignment of MALDI-TOF/TOF mass spectra of *N. sumatrana* venom peptides to protein families

The mass spectra were acquired using Data Explorer™ 4.0 software (Applied Biosystem®) and exported to the MS/MS ion search MASCOT peptide mass fingerprinting(http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF) to search against all non-redundant NCBI database (NCBInr), with taxonomy set to Serpentes (taxid:8570). To further confirm the assignment, the matched peptide sequences were subjected to manual NCBI Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with highest significant BLAST scores ($p < 0.05$) with lowest E-value.

5.2.7 Shotgun LC-MS/MS of the *N. sumatrana* venom

One milligram of venom sample was reduced, alkylated and digested with sequencing grade trypsin according to iTRAQ™ protocol (Applied Biosystems®). Peptides were analyzed by LC-MALDI using the UltiMate® 3000 Nano LC systems (Dionex) equipped with a ProBot robotic spotter (LC Packings) coupled to a AB SCIEX PepMap100 C_{18} nanocolumn (3 µm, 100 Å) and separated with gradient of 10-45% (v/v) acetonitrile containing 0.1% (v/v) TFA over 165 min. Mass spectra were acquired and analyzed for protein identification using MASCOT peptide mass fingerprinting

(http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF) with Uniprot Knowledge based (UniProKb) (released 21 Sep 2011) and taxonomy set to Serpentes (taxid: 8570).

5.2.8 *De novo* sequence analysis

For the peptides that were present in the *N. sumatrana* venom but are not assigned or well presented in database, *de novo* sequencing analysis was performed using high quality MS/MS spectra.

Naja sumatrana crude venom was resuspended in 100 ml of 0.1% (v/v) TFA in 2% (v/v) acetonitrile. From this solution 2 µl was added to 18 µl of 0.1% (v/v) TFA in 2% (v/v) acetonitrile. 10 µl was injected onto the UltiMate® 3000 Nano LC systems (Dionex, C_{18} PepMap100, 3 µm) and separated with a gradient of 10-40% (v/v) acetonitrile (containing 0.1% (v/v) TFA) with spotting using a ProBot (LC Packings) robotic spotter over 180 min. Peptides were analysed by MALDI-TOF/TOF mass spectrometry using a AB SCIEX TOF/TOF™ 5800 System (AB Sciex). MS/MS spectra were analyzed using PEAKS Studio Version 4.5 SP2 (Bioinformatics Solutions Inc).

All *de novo* peptide sequences were searched against NCBI Blast (http://blast.ncbi.nlm.nih.gov, non-redundant protein sequence (nr) database) with taxonomy set to Serpentes (taxid: 8570). The *de novo* peptides were first putatively assigned to a protein family using procedures described by Bringans *et al*., (2008), essentially by matching to those with highest BLAST score and lowest E-value. The accession numbers of the significant scored proteins were also checked to confirm the assignment.

Shotgun LC-MS/MS and *de novo* sequencing were conducted at Proteomics International, Perth, Australia. Data analysis and NCBI BLAST search of the *de novo* peptide sequences were completed in the University of Malaya, Kuala Lumpur, Malaysia.

5.2.9 Isolation of high molecular weight protein (HMW protein) from *N. sumatrana* venom by Sephadex® G-50 gel filtration column

Fifty milligram of lyophilized *N. sumatrana* venom was dissolved in 1 ml of 50 mM sodium phosphate buffer (pH 7.0). It was applied into Sephadex® G-50 gel filtration column (10 mm x 300 mm) that had been equilibrated with the same buffer and protein fractions were eluted at flow rate of 30 ml/h. The eluting proteins were detected by absorbance at 280 nm. Each 0.5 ml of eluting proteins was collected manually. The molecular weight of the fractions was assessed by SDS-PAGE as described in Chapter 3 (Section 3.2.2).

5.2.10 2-D Clean-Up

High molecular weight proteins were pooled and cleaned up by 2-D Clean-Up Kit according to manufacturer protocol (GE Healthcare). 2-D Clean-Up Kit consists of precipitant (15 ml), co-precipitant (17.5 ml), wash addictive (250 µl) and wash buffer (50 ml). Precipitant (300 µl) was added to 100 µl of protein sample and mixed thoroughly by vortexing. The mixture was then incubated in ice (4-5°C) for 15 min. Later, 300 µl of co-precipitant was added to the mixture and mixed thoroughly. The sample mixture was centrifuged at 12,000 xg for 5 min. The supernatant was discarded. Without dispersing the precipitate, 40 µl of co-precipitant was added and incubated in ice for 5 min. The sample mixture was centrifuged at 12,000 xg for 5 min and supernatant was discarded. Twenty-five microliter of UltraPure™ water was added and mixed by vortexing for 5-10 s. One milliliter of wash buffer (pre-chilled for at least 1 h at -20°C) and 5 µl of wash additive was added and mixed by vortexing to disperse the protein pellet. The mixture was incubated in -20°C for 30 min. The tube was centrifuged at 12,000 xg for 5 min and supernatant was discarded. The protein pellet was

allowed to air-dry for 3 min. The cleaned-up protein sample was dissolved in lysis buffer (7 M Urea, 2 M Thiourea, 4% (w/v) CHAPS) before protein quantitation.

Protein content was quantitated by 2D-Quant Kit (GE Healthcare) as described in Chapter 3 (section 3.2.1.2).

5.2.11 Two-dimensional electrophoresis (2DE) of high molecular weight (HMW) protein fractions from *N. sumatrana* venom

5.2.11.1 First dimension electrophoresis (Isoelectric focusing)

Rehydration solution was prepared from 7 M Urea, 2 M Thiourea, 4% (w/v) CHAPS (*i.e.* lysis buffer). Prior to rehydration, 0.002% (w/v) Orange G, 40 mM DTT and 0.5% (v/v) IPG buffer (Linear, pH 3-10) was added.

Approximately 200 µg of pooled HMW protein fractions dissolved in rehydration solution was loaded on IPG strips (13 cm, pH 3-10, linear) and allowed to rehydrate for 18 h using reswelling tray in IPGbox (GE Healthcare). After rehydration completed, the IPG strips were transferred onto a ceramic manifold with the gel side faced up. The IPG strips were placed with the anodic (+) end of the strips aligned with the appropriate mark etched, while the other end of the strips were aligned on the cathode (-) side. Wet paper wicks were placed to overlap the gel side of the IPG strips. Electrodes were assembled on the both anode and cathode platform. PlusOne DryStrip Cover fluid was added to cover the IPG strips in order to avoid crystallization of urea during isoelectric focusing.

Isoelectric focusing was carried out using Ettan™ IPGphor™ 3 Isoelectric Focusing System as described by the manufacturer (GE Healthcare).

Isoelectric focusing was carried out at 20°C using at five steps focusing program: 300 V for 30 min, 500 V for 1 h, 500-1000 V for 1 h, 1000-8000 V for 2.5 h and 8000 V for 1 h. After isoelectric focusing completed, the IPG strips were kept in -80°C before second dimension electrophoresis.

5.2.11.2 Second dimension electrophoresis (SDS-PAGE)

Electrophoresis reagents were prepared as follows:

SDS-PAGE electrophoresis buffer (10X): 0.25 M Tris, 1.92 M glycine, 1% (w/v) SDS, pH 8.3. SDS-PAGE electrophoresis buffer (1X and 2X) were prepared by diluting the 10X SDS-PAGE electrophoresis buffer.

Agarose sealing solution: 0.5% (w/v) agarose and 0.002% (w/v) Orange G was added into 1X SDS-PAGE electrophoresis buffer. The agarose sealing solution was heated to dissolve agarose.

SDS equilibration buffer: 50 mM Tris-HCl (pH 8.8), 6 M Urea, 30% (v/v) glycerol, 2% (w/v) SDS, 0.002% (w/v) Orange G. Prior to equilibration, 50 mg Dithiothreitol (DTT) and 125 mg Iodoacetamide (IAA) were added.

All the reagents for SDS-Polyacrylamide resolving gel were prepared as described in Chapter 3 (Section 3.2.2).

15% resolving gel (reducing) was prepared by mixing 50 ml of solution A, 25 ml of solution B, 1 ml of 1% (w/v) SDS, and 23 ml of UltraPure^{TM} water; 1 ml of 1% (w/v) APS and 33 µl of TEMED (N,N,N'N'-Tetramethylethylenediamine) was added prior to cast the gel.

SDS-PAGE was carried out using SE 600 Ruby Standard Dual Cooled Vertical Unit (GE Healthcare). The glass plates (160 mm x 180 mm) were assembled

on gel caster according to instructions by manufacturer (GE Healthcare). 50 ml of 15% resolving gel was loaded into the space between glass plates, 1 ml of 1X SDS-PAGE electrophoresis buffer was added to overlay the resolving gel. The resolving gel was allowed to polymerize for 1 h at room temperature.

Prior to SDS-PAGE, the IPG strips were rinsed in 1X SDS-PAGE electrophoresis buffer and reduced in 5 ml of equilibration buffer containing 50 mg of DTT for 15 min, and alkylated in 5 ml of equilibration buffer containing 125 mg IAA for 15 min. The equilibrated IPG strips were applied into the 15% resolving gel (1.5 mm thick gel). Agarose sealing solution was added to overlay the gel strips. The glass plates were assembled with upper chamber into an electrophoresis tank. Upper chamber was filled with 2X SDS-PAGE electrophoresis buffer, while lower chamber was filled with 1X SDS-PAGE electrophoresis buffer. Electrophoresis was carried out at two steps at constant volt: 15 mA/ gel, 9 V for 20 min and 50 mA/gel, 230 V for 5 h.

Upon completion of electrophoresis, the plates were opened, and the gel was gently removed, and washed in UltraPure™ water for 5 min before stained with silver stain (Pierce® Silver Stain for Mass Spectrometry) as described in Chapter 3 (section 3.2.2.2). The protein spots viewed on silver-stained gels using Image Master 2D Platinum 7.0 software (GE Healthcare), were excised, reduced, alkylated, and digested with Trypsin Gold as described in section 5.2.2 and analyzed by MALDI-TOF/TOF as described in section 5.2.3. Protein spots from at least 5 gels were pooled for MALDI-TOF/TOF analysis.

5.3 RESULTS

5.3.1 Proteomic characterization of the *N. sumatrana* venom proteins isolated and purified from ion exchange and reverse-phase chromatography

Resource® S ion exchange chromatography of *N. sumatrana* venom yielded 11 fractions, as shown in Figure 5.1. Fractions 3-6 exhibited phospholipase A_2 activity. All the fractions were subjected to reverse-phase C_{18} chromatography (Figure 5.2). Table 5.1 shows the identity of the fractions purified by reverse-phase HPLC as analyzed by MALDI-TOF/TOF. There are few minor fractions which consist only trace amount of the proteins and were not further analyzed. Reverse-phase HPLC of fraction 1, the acidic protein fraction that constitutes 4% of total venom proteins yielded multiple fractions that were poorly resolved (Figure 5.2A) and were not further analyzed.

Fraction 2 yielded one major fraction and two minor fractions in reverse-phase HPLC, MALDI-TOF/TOF analysis showed that the major fraction was identified as long neurotoxin (accession number: O42257, termed *Nsumatrana* neurotoxin 1). Fraction 3+4 yielded one major and one minor fraction, both were identified to be acidic phospholipase A_2 (Fraction 3+4a, the major fraction, was termed *Nsumatrana* PLA_2-1). Fraction 5 yielded six fractions but only five were identified (5a, b, c, d, e) as various isomers of neutral phospholipase A_2 and acidic phospholipase A_2. Fraction 6, the major phospholipase A_2 fraction of the venom, yielded one major fraction in reverse-phase HPLC and was identified as a neutral phospholipase A_2 (termed *Nsumatrana* PLA_2-2). There were some minor fractions which were not identified. Fraction 7+8 yielded one major (b) and one minor (a) fraction, however only the major fraction was identified as a short neurotoxin (termed *Nsumatrana* neurotoxin 2). Fraction 9 was resolved

into multiple fractions by reverse-phase HPLC. The first two fractions (a and b) were both identified to be neurotoxins, while the other two more hydrophobic fractions (c and d) were identified to be cardiotoxins. Fraction 10 was identified as a cardiotoxin (termed *Nsumatrana* CTX-1), while fraction 11, the most basic venom protein fraction yielded a minor and a major fraction that were both identified as cardiotoxin, the minor cardiotoxin was termed *Nsumatrana* CTX-2 and the major cardiotoxin as *Nsumatrana* CTX-3.

A combination of ion exchange and reverse-phase HPLC thus identified 17 proteins including 8 phospholipases A_2, 4 neurotoxins and 5 cardiotoxins. However, only 7 of the proteins exist in significant quantity: the long neurotoxin *Nsumatrana* NTX-1 (8.7%), the acidic phospholipase A_2 *Nsumatrana* PLA_2-1 (4.6%), the neutral phospholipase A_2 *Nsumatrana* PLA_2-2 (26.0%), the short neurotoxin *Nsumatrana* NTX-2 (3.5%), and the three cardiotoxins: *Nsumatrana* CTX-1 (15.4%), CTX-2 (1.6%) and CTX-3 (27.2%). These major toxins constitute approximately 87% of total venom proteins. The other 10 proteins exist in minor amounts. The major and minor proteins constitute approximately 89% of the venom proteins and represents three types of toxins phospholipase A_2 (32%), neurotoxin (12.3%) and cardiotoxin (45%). Figure 5.3 shows the SDS-PAGE of 6 of the most abundant proteins. All isolated toxins were homogenous as shown in SDS-PAGE. The molecular weight of the toxins were at expected value, except for the short neurotoxin 3, which exhibited a molecular weight of ~12 kDa, while typical cobra venom short neurotoxin has a molecular weight of 7 kDa (Tan, 1983).

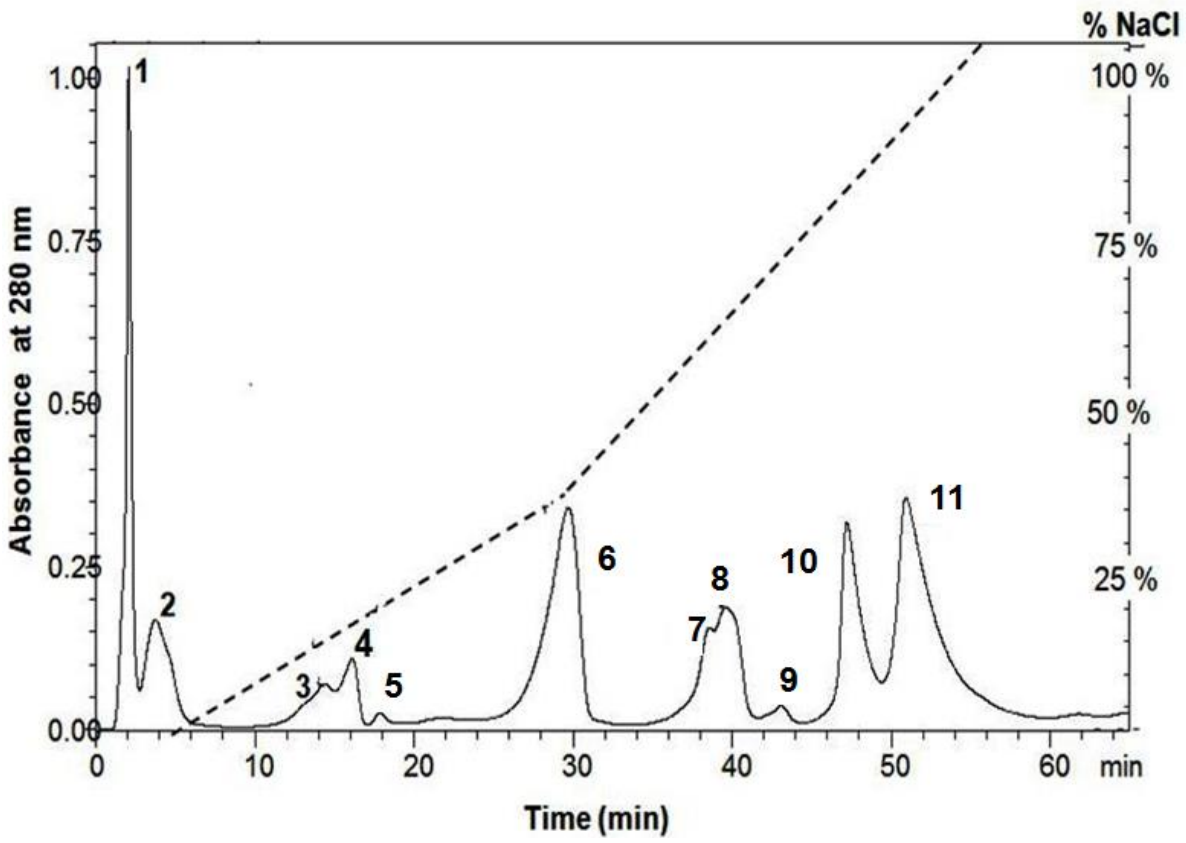

Figure 5.1: Resource® S ion exchange chromatography of *N. sumatrana* venom.

Ten milligrams of *N. sumatrana* venom was injected into the column equilibrated with 20 mM MES, pH 6.0, and eluted by a linear gradient (of 0-30% of 0.8 M NaCl from 5 to 30 min, followed by 30-100% from 30 to 55 min). Flow rate was 1 ml/min.

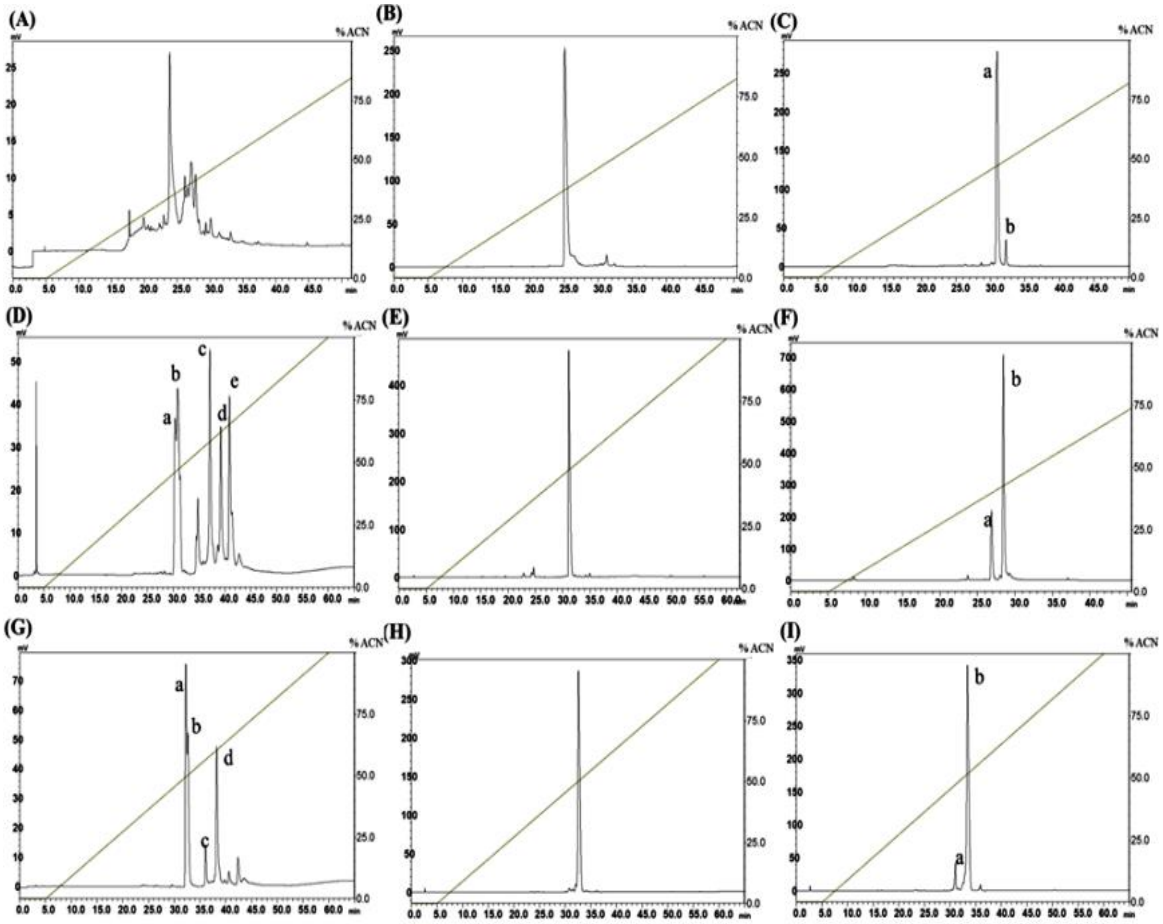

Figure 5.2: C_{18} reverse-phase chromatography of *N. sumatrana* venom fractions isolated from Resource® S ion exchange chromatography.

The isolated fractions were injected into the column equilibrated with 0.1% (v/v) TFA and eluted by a linear gradient 0-100% (v/v) acetonitrile containing 0.1% (v/v) TFA for 60 min. Flow rate was 1 ml/min. Straight line: elution gradient.

(A) Fraction 1; (B) Fraction 2; (C) Fraction 3+4; (D) Fraction 5; (E) Fraction 6; (F) Fraction 7+8; (G) Fraction 9; (H) Fraction 10; (I) Fraction 11.

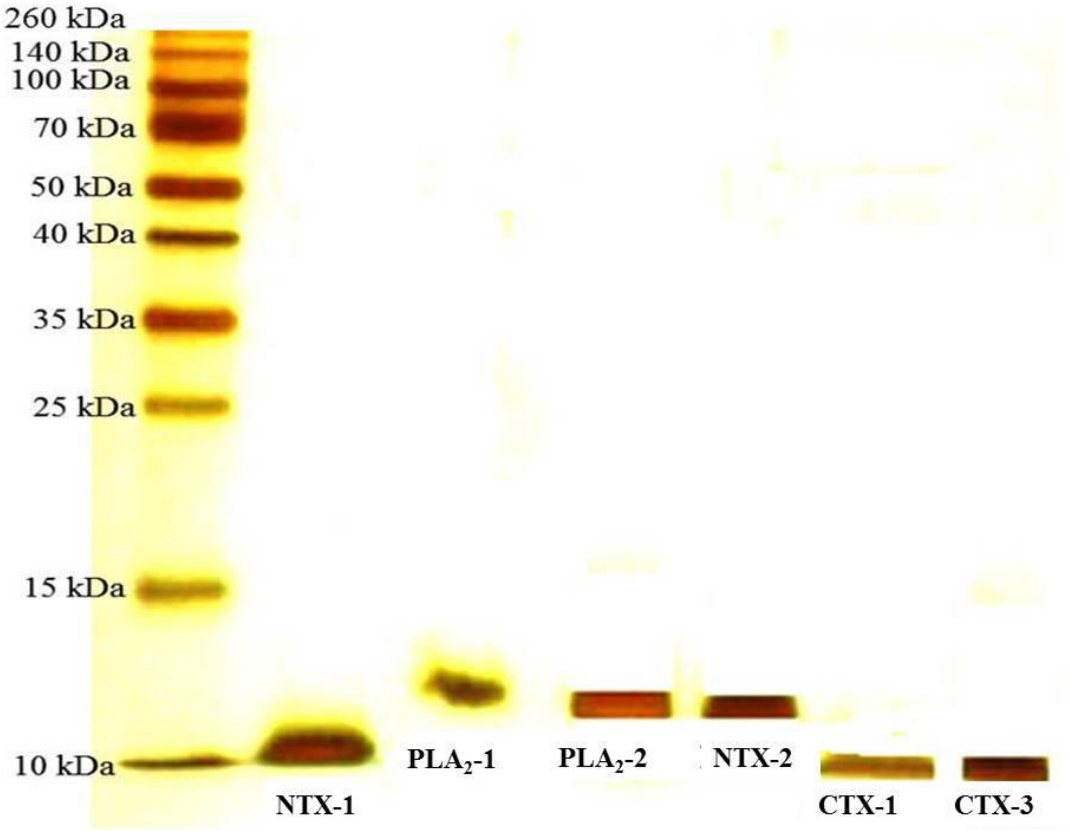

Figure 5.3: SDS-PAGE of major fractions isolated and purified from *N. sumatrana* venom.

The major protein fractions were loaded into 15% (reducing) SDS- Polyacrylamide gel and visualized by silver staining. Left lane: molecular weight standards.

The six toxins were: NTX-1 (*Nsumatrana* NTX-1, a long neurotoxin); PLA_2-1 (*Nsumatrana* PLA_2-1, an acidic phospholipase A_2); PLA_2-2 (*Nsumatrana* PLA_2-2, a neutral phospholipase A_2); NTX-2 (*Nsumatrana* NTX-2, a short neurotoxin); CTX-1 (*Nsumatrana* CTX-1) and CTX-3 (*Nsumatrana* CTX-3)

Table 5.1: Assignment of ion exchange isolated and reverse-phase purified fractions of *N. sumatrana* venom to protein families by MALDI-TOF-TOF.

Fraction No.		Protein/protein families	MS/MS matched peptide sequence	Accession No.	Protein score	% sequence coverage	% of venom protein
2		Long Neurotoxin 7 (termed *Nsumatrana* NTX-1)	TWCDGF CSSR CFITPDVTSTDCPNGHVCYTK VELCCAATCPTVKPGVDIQCSTD	O42257	73	43	8.7
3+4	a	Phospholipase A_2, acidic D (termed *Nsumatrana* PLA_2-1)	NMIQCTVPNR GGSGTPVDDLDR CCQVHDNCYGEAEK CWPYFK TYSYECSQGTLTCK GGNDACAAAVCDCDR LAAICFAGAPYNDNNYNIDLK	Q9I900	566	66	4.6
	b	Phospholipase A_2, acidic C	SWWDFADYGCYCGR GGSGTPVDDLDRCCQVHDNCYGEAEK CCQVHDNCYGEAEK GGNNACAAAVCDCDR	Q92086	174	51	0.3
5	a	Phospholipase A_2, neutral A	NMIQCTVPNR GGSGTPVDDLDR CCQIHDNCYNEAEK CWPYFK TYSYECSQGTLTCK GGNNACAAAVCDCDR LAAICFAGAPYNDNNYNIDLK	Q92084	594	63	0.4

Table 5.1 (continued)

	b	Phospholipase A_2, acidic D	NMIQCTVPNR CCQVHDNCYGEAEK CWPYFK TYSYECSQGTLTCK GGNDACAAAVCDCDR	Q9I900	402	48	0.04
	c	Phospholipase A_2, neutral A	NMIQCTVPNR SWWHFADYGCYCGR CWPYFK TYSYECSQGTLTCK GGNNACAAAVCDCDR LAAICFAGAPYNDNNYNIDLK	Q92084	118	54	0.2
	d	Phospholipase A_2, neutral A	NMIQCTVPNR SWWHFADYGCYCGR GGSGTPVDDLDR CCQIHDNCYNEAEK CWPYFK LAAICFAGA PYNDNNYNIDLK	Q92084	121	52	0.01
	e	Phospholipase A_2, neutral A	NMIQCTVPNR GGSGTPVDDLDR CCQIHDNCYNEAEK CWPYFK TYSYECSQGTLTCK GGNNACAAAVCDCDR LAAICFAGAPYNDNNYNIDLK	Q92084	594	63	0.2
6		Phospholipase A_2 , neutral A (termed *Nsumatrana* PLA_2-2)	SWWHFADYGCYCGR GGSGTPVDDLDR CCQIHDNCYNEAEK CWPYFK TYSYECSQGTLTCK GGNNACAAAVCDCDR	Q92084	527	51	26.0

Table 5.1 (continued)

7+8	a	No significant hit					
	b	Neurotoxin 3 short (termed *Nsumatrana* NTX-2)	LECHDQQSSQTPTTTGCSGGETNCYK NGIEINCCTTDR	Q9PSN6	82	74	3.5
9	a	Neurotoxin, partial	LECHDQQSSQAPTTTGCSGGETNCYK GCGCPSVK KGIEINCCTTDR	AAA66026	143	74	0.1
	b	Neurotoxin 3 Short	LECHDQQSSQTPTTTGCSGGETNCYK GCGCPSVK NGIEINCCTTDR	Q9PSN6	179	75	0.02
	c	Cardiotoxin 2b	LVPLFYK MYMVATPK RGCI DVCPK YVCCNTDR	Q9PST3	166	39	0.1
	d	Cardiotoxin 4b	LVPLFYK MYMVAMPK RGCIDVCPK GCIDVCPK YVCCNTDR	O73856	181	39	0.1
10		Cardiotoxin 5a precursor (termed *Nsumatrana* CTX-1)	LVPLFYK MFMVSNLTVPVK MFMVSNLTVPVKR GCIDVCPK YVCCNTDR	O73857	269	44	15.4
11	a	Cardiotoxin 6 precursor (termed *Nsumatrana*	MFMVSNK RGCIDVCPK GCIDVCPK YVCCNTDR	O73858	139	38	1.6

Table 5.1 (continued)

		CTX-2)					
	b	Cardiotoxin 3 precursor (termed *Nsumatrana* CTX-3)	LVPLFYK MFMVATPK RGCIDVCPK GCIDVCPK YVCCNTDR	P60302	233	39	27.2

Protein scores greater than 67 are significant ($p < 0.05$)

The mass spectra acquired were searched against all non-redundant NCBI database with taxonomy set to Serpentes (taxid: 8570).

5.3.2 Shotgun LC-MS/MS proteomic characterization of *N. sumatrana* venom

Figure 5.4 shows the elution profile of the trypsin digested peptides from *N. sumatrana* venom by reverse-phase C_{18} LC, monitored at 214 nm. Using the shotgun approach and enrichment with *de novo* sequencing method, a total of 50 individual proteins were identified from *N. sumatrana* venom (Table 5.2 and 5.3). For the shotgun analysis, 75% of the proteins were identified by more than one matched peptide.

The proteins identified by shotgun LC-MS/MS are phospholipase A_2 and related proteins (18 proteins), and three-finger toxins (including neurotoxins, cardiotoxins and cytotoxins) and related proteins (19 proteins), which together constitute 74% of the proteins identified. The other proteins identified include the high molecular weight proteins thaicobrin, aminopeptidase, Cysteine rich secreting protein(CRISP 2 include kaouthin-1), cobra venom factor, complement depleting factors (2), zinc metalloproteinase-disintegrin (cobrin), venom nerve growth factor, NADH dehydrogenases (2), cobra serum albumin and the low molecular weight proteins including natriuretic peptide, cytotoxin and a neurotoxin homolog (Figure 5.5).

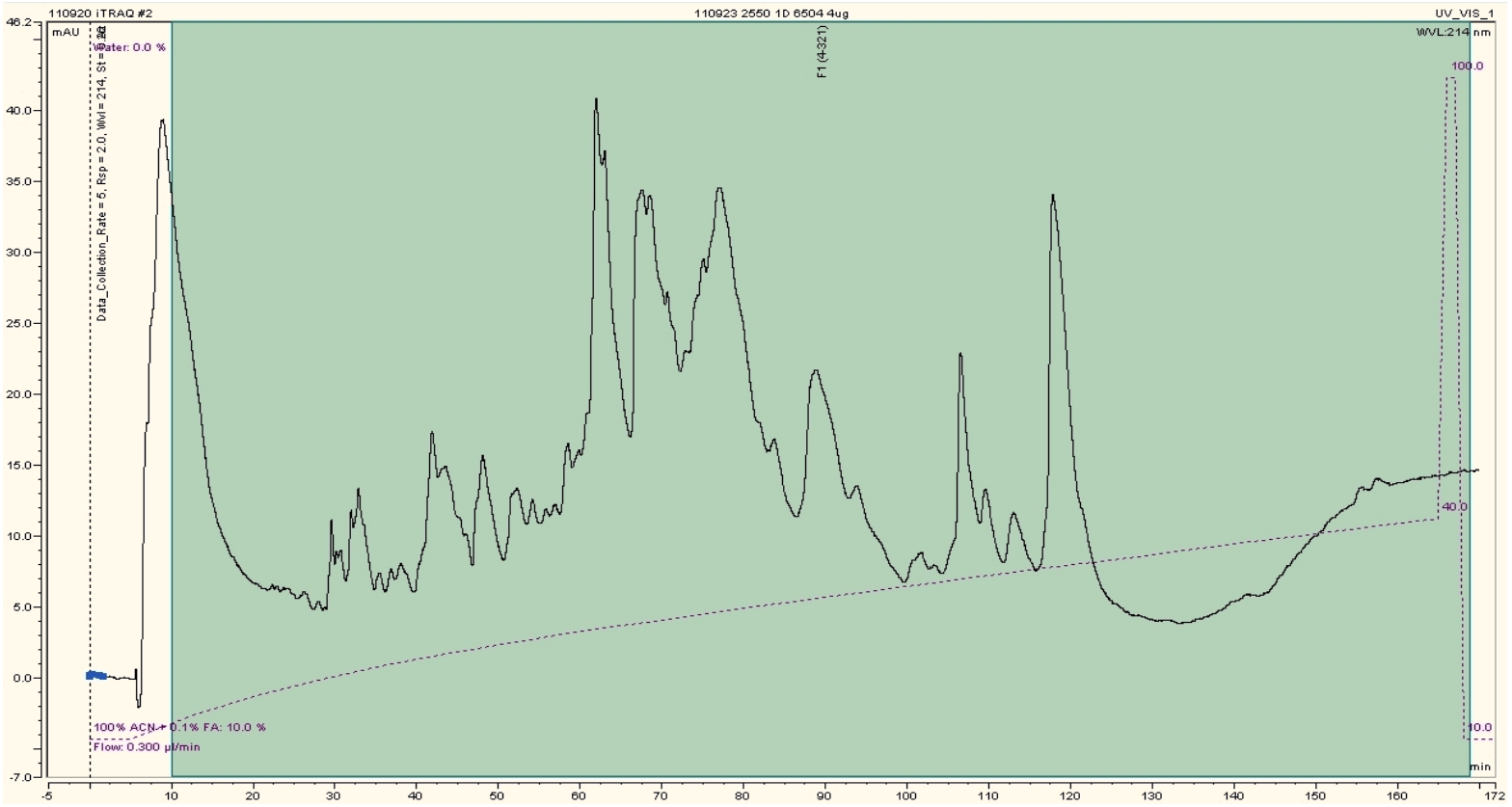

Figure 5.4: Shotgun LC-MS/MS chromatography of trypsin digested *N. sumatrana* venom peptides.

The trypsin digested peptides were loaded into C_{18} PepMap100 reverse-phase column connected to 5800 MALDI-TOF/TOF Analyzer and resolved by a gradient of 10-45% acetonitrile (0.1% (v/v) TFA) over 165 min. Flow rate was 0.3 µl/min. The eluted peptides were monitored by absorbance at 214 nm.

Table 5.2: Assignment of shotgun LC-MS/MS for trypsin digested *N. sumatrana* venom proteins to protein families.

Protein/protein families	MS/MS matched peptide sequence	Accession No.	No of match peptide	Protein score	Query coverage (%)
Thaicobrin	SPPGNWQK SGKHFFEVK FDGSPCVLGSPGFR TVENVGVSQVAPDNPER KTVENVGVSQVAPDNPER ADVTFDSNTAFESLVVSPDKK	P82885	6	289	63.9
Cytotoxin KJ C-3	LVPLFYK YVCCNTDR YVCCNTDRCN MFMVSDLTVPVKR MFMVSDLTVPVKR	P86541	5	248	50
Cardiotoxin 4N	LVPLFYK YVCCNTDR YVCCNTDRCN MFMVSNLTVPVKR	Q9W6W9	4	241	37
Cobra venom factor	VNDDYLIWGSR GDNLIQMPGAAMK QLDIFVHDFPR ASVQEALWSDGVR AVPFVIVPLEQGLHDVEIK	Q91132	5	169	4.1
Cardiotoxin OS/ Cardiotoxin 3	LVPLFYK MFMVATPK YVCCNTDR	Q02454	3	114	30.9

Table 5.2 (continued)

Phospholipase A_2 isoenzyme CM-II	NLYQFK NMIQCTVPNR NLYQFKNMIQCTVPNR	P07037	3	82	13.4
Phospholipase A_2 isoenzyme DE-II	NLYQFK NMIQCTVPNR GGSGTPVDDLDR NMIQCTVPNR NLYQFKNMIQCTVPNR	P00600	5	81	23.5
Phospholipase A_2 acidic D	CWPYFK NMIQCTVPNR GGSGTPVDDLDR NMIQCTVPNR SWWDFADYGCYCGR LAAICFAGAPYNDNNYNIDLK	Q9I900	6	78	43.2
Cardiotoxin 4b	LVPLFYK MYMVAMPK YVCCNTDR	O73856	3	75	30.9
Cardiotoxin 4a	LVPLFYK MFMVAMPK YVCCNTDRCN	O93474	3	75	30
Complement depleting factor	VFSMDHNMR QLDIFVHDFPR LDDKVPGTEIETK	ABN72543	3	70	3.3
Cardiotoxin-2b	LVPLFYK YVCCNTDRCN	Q9PST3	2	69	39
Phospholipase A_2,	GGSGTPVDDLDR	Q6T179	5	62	59.5

Table 5.2 (continued)

isoform 4	GRNNACAATVCDCDR SNRPMPLNIYQFK SWWDFADYGCYCGR LAAICFAGAPYNDNNYNIDLK				
Cytotoxin-10	MYMVATP LVPLFYK	P86541	2	57	46.7
Phospholipase A_2, acidic C	CWPYFK NMVQCTVPNR GGSGTPVDDLDR SWWDFADYGCYCGR LAAICFAGAPYNDNNYNIDLK	Q92086	5	60	43.2
Phospholipase A_2, isoform 3	GGSGTPVDDLDR SNRPMPLNLYQFK	P60045	2	55	35
Zinc metalloproteinase disintegrin Cobrin	DPSYGMVEPGTK CGDGMVCSNRQCVDVK	Q9PVK7	2	54	4.7
Phospholipase A_2	NLYQFK GGSGTPVDDLDR SWWDFADYGCYCGR LAAICFAGAPYNDNNYNIDLK	Q65ZF5	4	46	15.3
Cytotoxin-3	MYMVSTLTVPVKR	P01459	1	51	21.7
Phospholipase A_2	NLYQFK NMIQCANKGSR	P20258	2	46	14.5
Phospholipase A_2	NLYQFK GGKGTPIDDLDR	P14556	2	46	15.3
Phospholipase A_2 isozyme E	NLYQFK GGSGTPVDDLDR	P25498	2	45	43

Table 5.2 (continued)

Phospholipase A_2 isozyme 1	NYNIDLK DFVCNCDR CYDTAEKVHGCWPK CCQVHDKCYDTAEK	B3EXG8	4	42	25.9
Kaouthin-1	EIVDLHNSLR	B5THG8	1	34	9.6
Cysteine-rich secretory protein 1	NVDFNSESTR EIVDLHNSLR	P86543	2	34	60.6
Neurotoxin homolog	ISLADGNDVR RGCTFTCPELRPTGK	Q9DEQ3	2	33	29.1
PLA_2-2 precursor	APYNTKNFMILDK	ABK63564.1	1	27	8.6
Natriuretic peptide	GSSCFGQK IGSMSGMGCR ILEYLRPDSK	D9IX97	3	27	18.8
Putative three-finger toxin	WHMLAPGR RWHMLAPGR	F5CPE5	2	27	10.5
Alpha- neurotoxin NTX 2	SWRDHR GIEINCCTTDR KGIEINCCTTDR	Q9YGJ5	3	26	21.7
Long neurotoxin 3	TWCDAWCSR	P25667	1	26	12.5
Three-finger toxin	YVCCNTDR	E2IU05	1	25	75
Aminopeptidase	MAKGFFISK AQIIDDAFNLAR LMFDRSEVYGPMQR EYLPWDTALDNLDYFR	B6EWW5	4	25	8.6
Phospholipase A_2 isoform 1	NMIQCTVPKR SWWDFADYGCYCGR LAAICFAGAPYNDNNYNIDLK	P60043	3	25	35.7
Phospholipase A_2	LAAICFAGAPYNDNNYNIDLK	Q5G291	1	25	37

Table 5.2 (continued)

isoform 5					
Short neurotoxin 1	GIELNCCTTDR KGIELNCCTTDR LECHNQQSSEPPTTTRCSGGETNCYK	P01431	3	24	61.3
Cobra serum albumin	SKPNISEEELAATILTFR	Q91134	1	23	2.9
Cobrotoxin	LECHNQQSSQTPTTK	P59275	1	23	24.6
Cytotoxin- 1	YECCNTDR	P01447	1	22	13.3

The mass spectra acquired were searched against all non-redundant NCBI database with taxonomy set to Serpentes (taxid: 8570). Protein families obtained from shot-gun LC-MS/MS were sorted based on protein scores.

Table 5.3: Assignment of the trypsin digested peptides of *N. sumatrana* venom proteins by *de novo* sequence analysis.

***De novo* deduced peptide sequence**	***m/z***	**Protein family (Taxonomy = Serpentes)**	**Accession No.**	**% sequence coverage**
AAFLECCR	1004.36	Complement depleting factor	ABN72543.1	0.5%
DGHVAVMYTK	1120.47	Long neurotoxin	P34074	14%
MLVQCGSAAGR	1138.68	Phospholipase A_2-4 precursor	ABK63573.1	7.5%
LMLQCTVNPR	1220.53	Phospholipase A_2 3	P86542.1	67%
DMVECCSTDR	1250.40	weak neurotoxin 8 precursor/ weak neurotoxin 6 precursor	AAL87466.1/ AAL87465.1	12%
LYPDSFCTSAR	1305.56	Acanthophin-D (Postsynaptic neurotoxin)	P34073.1	15%
DTNNQTVNPR	1307.55	NADH Dehydrogenase subunit 2	AAW63026.1	3%
DTNNQTVGGPR	1307.55	NADH Dehydrogenase subunit 2	AAW63026.1	3%
LTHNFMQMFGK	1502.59	β-bungarotoxin phospholipase A_2 isotoxin	AAB33949	34%
DTNCNQTVNPR	1307.55	Venom nerve growth factor	P34128.1	4.5%
DDNACAAAVCD	1579.50	Phospholipase A_2, acidic	P15445.1	9%

The *de novo* deduced peptide sequences were subjected to NCBI blast (http://blast.ncbi.nlm.nih.gov Non-redundant protein sequence (nr) database) with taxonomy set to Serpentes (taxid: 8570). The *de novo* peptides were first putatively assigned to a protein family by matching to those with highest BLAST score and lowest E-value.

m/z: mass to charge ratio which is a pattern representing the distribution of ions by mass, and usually on x-axis of mass spectrum.

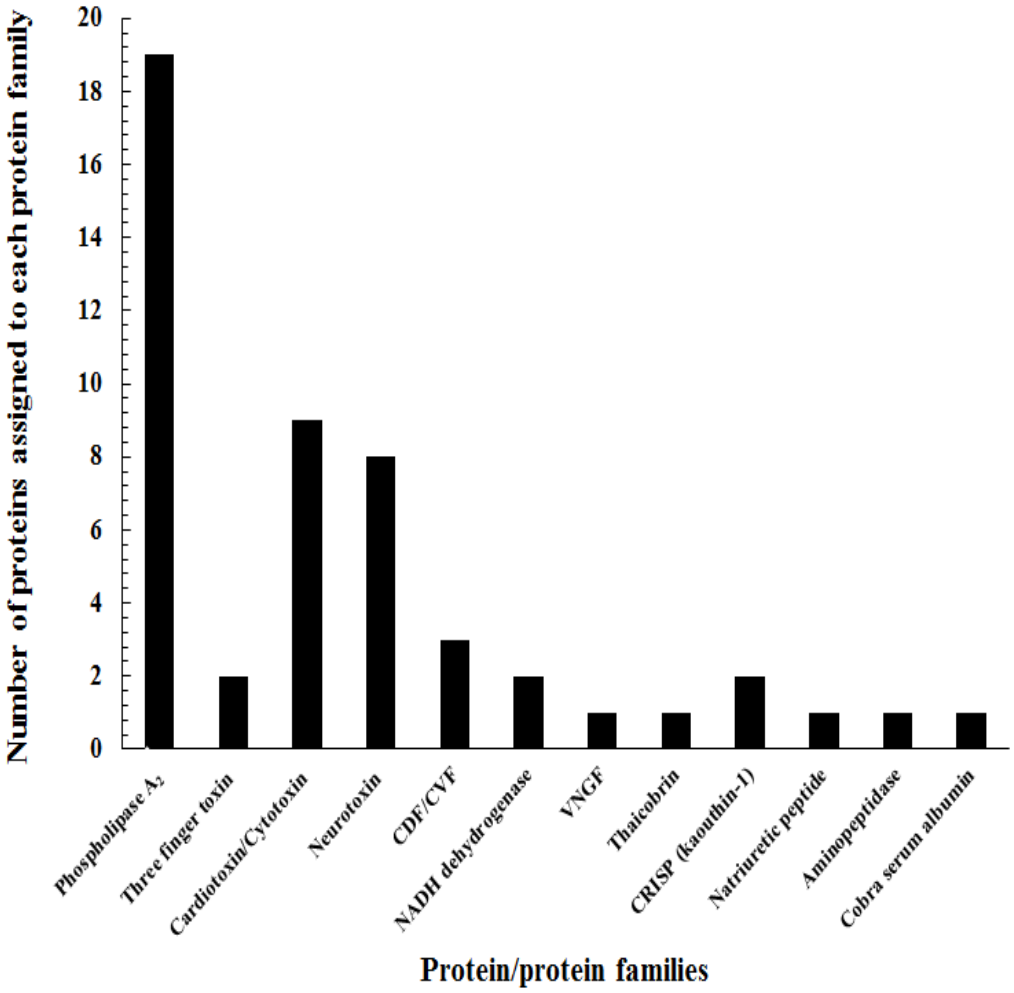

Figure 5.5: Types of proteins determined by shotgun LC-MS/MS approach coupled with *de novo* peptide sequencing.

Approximately 74% of the proteins represent phospholipase A_2 and three-finger toxins related proteins. CDF: complement depleting factor; CVF: cobra venom factor; VNGF: venom nerve growth factor; CRISP: Cysteine rich secretory protein

5.3.3 Isolation of the high molecular weight proteins of *N. sumatrana* venom

Sephadex® G-50 gel filtration chromatography of *N. sumatrana* venom yielded four major fractions (Figure 5.6). Fraction 1 consists of high molecular weight proteins whereas fraction 2 consists mainly of low molecular weight polypeptides. The protein composition of the tubes constituting fraction 1 was examined by SDS-PAGE (15% reducing gel) (Fig. Insert). The electrophoretic patterns indicated that fractions 19 to 24 consist mainly of the high molecular weight proteins, whereas in fractions 29-30 contains the bulk of the proteins which were of low molecular weight, mainly the polypeptide toxins and phospholipases A_2. Fractions 19-24 were pooled and designated as high molecular weight venom protein fraction. SDS-PAGE, however, showed that there were some low molecular weight proteins also present in the pooled sample.

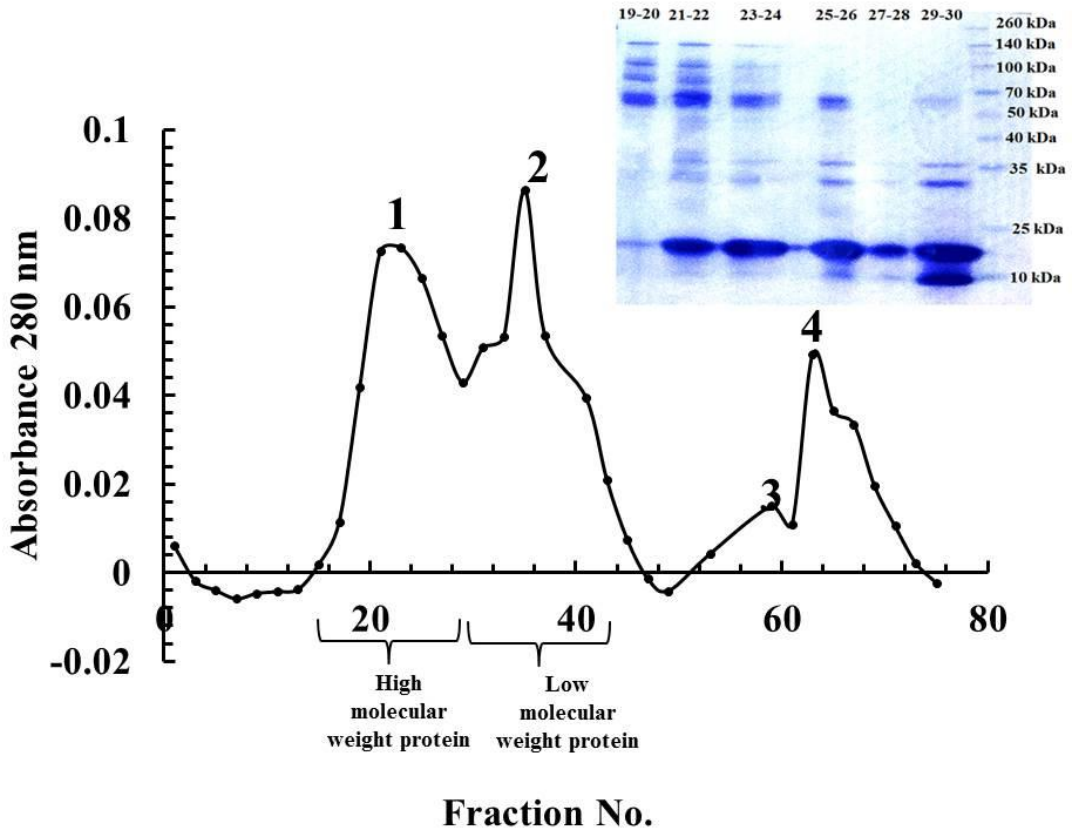

Figure 5.6: Sephadex® G-50 gel filtration chromatography of *N. sumatrana* venom. The venom (50 mg) was loaded to the column (10 mm x 300 mm) and eluted by 50 mM phosphate buffer (pH 7). Fractions of 0.5 ml were collected. The insert shows the SDS-PAGE (15%, reducing gel) of the fractions. The right lane shows the broad molecular mass protein marker. Fractions 19-24 were pooled and designated as the high molecular weight venom protein fractions.

5.3.4 Two dimensional electrophoresis of the high molecular weight protein fraction of *N. sumatrana* venom and identification of the proteins by MALDI-TOF/TOF

Figure 5.7 shows the 2DE pattern of the *N. sumatrana* venom high molecular weight protein fraction. A total of 31 clear and reproducible spots (n = 5 gels) were obtained after silver staining. MALDI-TOF/TOF analysis of the excised and treated spots revealed the identity of 27 proteins. Four of the spots failed to yield significant hit (Table 5.4). Of the 27 proteins, there were 6 phospholipases A_2, a long neurotoxin and a cardiotoxin. Five of the 6 phospholipases A_2 (spots 23-25, 27 and 28), the long neurotoxin (spot 30, corresponds to *Nsumatrana* NTX-1) and the cardiotoxin (spot 31, corresponds to *Nsumatrana* CTX-3) were also detected in the Resource® S ion exchange fractions mentioned in the earlier section (section 5.3.1). The presence of these low molecular weight proteins in the high molecular weight fraction was indicated by the presence of low molecular weight protein bands in the SDS-PAGE of fractions 21-24.

The other 19 high molecular weight proteins identified include: oligophrenin-1, complement depleting factor, cobra venom factors (2) and related proteins including cobra venom factor chains B (3), cobra venom factor chains G (2) and venom factor precursors (2), plastin-3, serum albumin precursor, phosphodiesterase, snake venom 5'-nucleotidase, hemorrhagic metalloproteinase-disintegrin (kaouthiagin), snake venom serine proteinase, Natrin (venom CRISP) chain A and venom nerve growth factor.

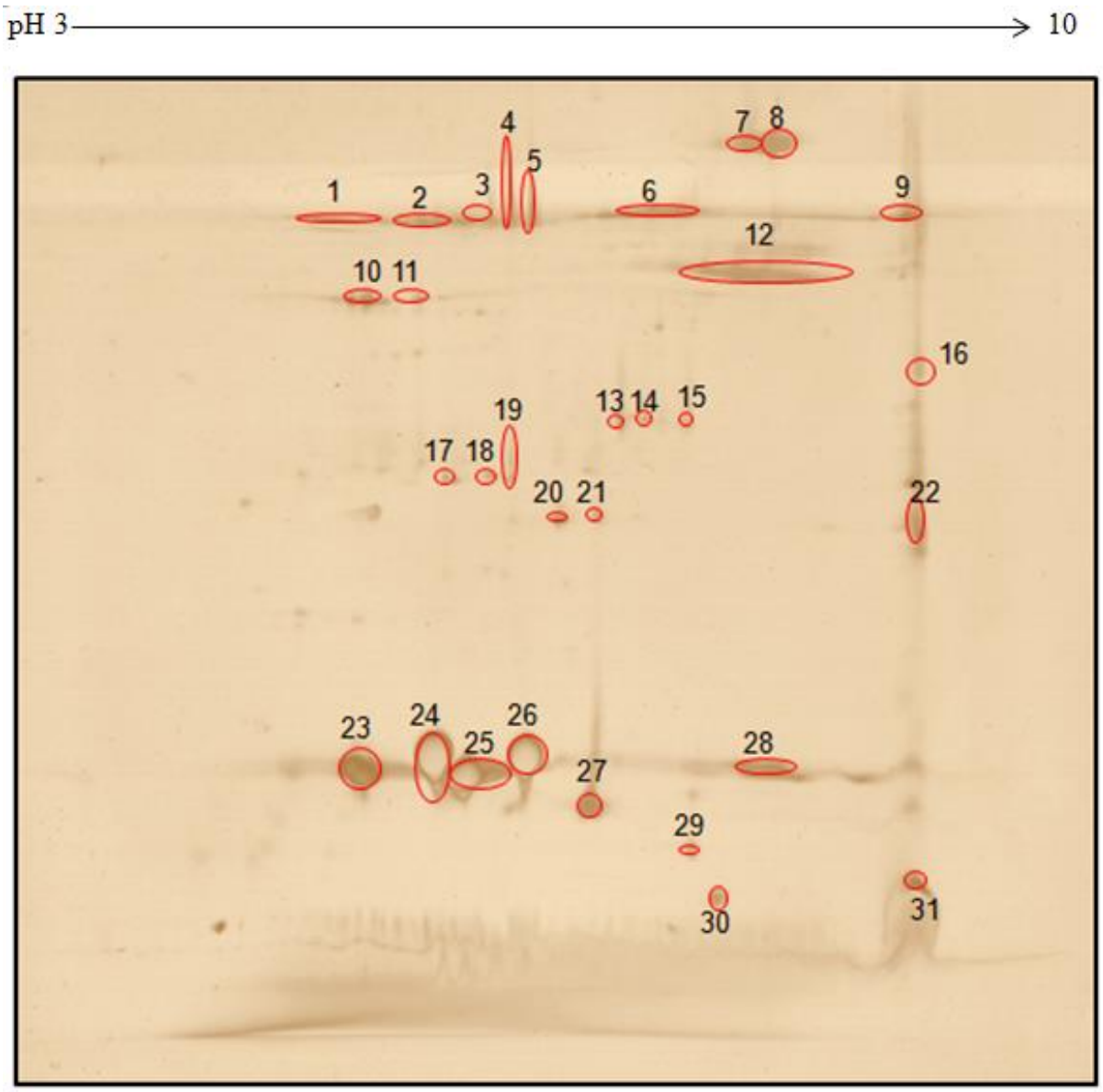

Figure 5.7: Two-dimensional gel electrophoresis of the high molecular weight protein fraction isolated from Sephadex® G-50 gel filtration chromatography.

The high molecular weight protein fraction (200 µg) was loaded on 13 cm IPG strips (pH 3-10, linear) for isoelectric focusing. The samples on the IPG strips were then reduced and alkylated in iodoacetamide, and then subjected to SDS-PAGE (15%, reducing gel). The gel was subsequently stained with Pierce® Silver Stain for Mass Spectrometry.

Table 5.4: Assignment of the two-dimensional spots of high molecular weight proteins (HMW proteins) isolated from *N. sumatrana* venom (Figure 5.7) to protein families by MALDI-TOF/TOF.

Spots	Protein families	MS/MS matched peptide sequences	Accession No.	% of sequence coverage	Protein score
1	Oligophrenin-1 [*Crotalus adamanteus*]	YYCQYEK TVGSNIQVQK LQEGGAKLSSK FFETASFRR MTPMEQKPGAK HLVHVCEHSR CSEQELERTNK STIEGYLYSQEK AQEDTVAAMMNIK MLRMTPMEQKPGAK LPGKSTIEGYLYSQEK	AFJ50840	27	44
2	No significant hit				
3	No significant hit				
4	Venom factor precursor	IWDTIEK QYVSQRK TEAVQQIK QQPDGVFK AADYLLKK GVDRYISK IQKPGAAMK VLMAASTER GAVIIYLDK EDELFLAR FAEANPVVR VPDTEIETK LKVVPEGMR IQKPGAAMKIK ASVRGQLASDGVK GGTDRNPQAKPR GIYTPGSPVRYR	AFJ49268	33	55
5	Venom factor precursor	CPQPENR SIPESLTR QYVSQRK VVPEGMRK TEAVQQIK VVPEGMRK VLMAASTER GAVIIYLDK VPDTEIETK LKVVPEGMR ASVRGQLASDGVK HFEVGFIQPGSVK FLYIDGNTDFHVSITARYL	AFJ49268	33	79

Table 5.4 (continued)

		YGK			
6	complement-depleting factor, partial [*Bungarus fasciatus*]	GELASDGVR VVPEGERK IQKPGAAMK GELASDGVRK VPETEIETK IKLEGDPGAR ISYENAFLAR VAQFQDQDLR SVAKCPQPANR ADHSYAAFPNR AAFLECCLYIK DIVTIIELDPSEK CAEETCSLLNQQK IEEKDGNDIYVMDVLE VIK	ABN72542	28	61
7	Cobra venom factor	QNQYVVVQVTGPQVR GIYTPGSPVLYR VFSMDHNTSK YEHSPENYTAYFDVRK KYVLPSFEVR SIPDSLTR IPIIDGDGK TNHGDLPR IKLEGDPGAR LEGDPGAR IWDTIEK CPQPANR AAEFQDQDLRK AKYIQEGDACK LDDRVPDTEIETK VLMAASTGR ICIGNVCR CQEALNLK	Q91132.1	35	136
8	Cobra venom factor	QLDIFVHDFPR QNQYVVVQVTGPQVR GIYTPGSPVLYR VFSMDHNTSK KYVLPSFEVR YVLPSFEVR IPIIDGDGK TNHGDLPR GDNLIQMPGAAMK IKLEGDPGAR IWDTIEK CPQPANR AKYIQEGDACK LDDRVPDTEIETK VLMAASTGR ICIGNVCR CQEALNLK	Q91132.1	37	156
9	Plastin-3 [*Crotalus*	YAVSMARR	AFJ50952	25	60

Table 5.4 (continued)

	adamanteus]	AVGDGIVLCK INNFSSEIK HSNAKYAVSMAR RAECMLQQADR MVNLSVPDTIDER TGDLSEEDKHSNAK HVIPMNPNTDDLFK MVMTVFACLMGRGLK EGEPEININMSGFNEK			
10	Serum albumin precursor - monocled cobra	MKWVIFISLLCLVSFAE VK YGINDCCAK ADPDRNECVLSHK RYPTALSVVILESTK FREIMEEQEYTCYNLK EIMEEQEYTCYNLK AALSQYVCEHK SPDLPPPSEEILKETEAC TTYTEQR ETEACTTYTEQR ELGDYFFTNEFLVK MMPQAPTSFLIELTEK WECISNLGPDLSFVPPT FNPK SKPNISEEELAATILTFR GQEMVEHLQNGPTTE	S59517	30	249
11	Phosphodiesterase [*Crotalus adamanteus*]	AATYFWPGSEVK	AFJ49200.1	35	68
12	Snake venom 5'-nucleotidase	VPTYVPLEK QAFEHSVHR QVPVVQAYAFGK VVSLNVLCTECR	BAG82601	30	105
13	Chain B of Cobra Venom Factor	QGTDENPR GVDRYISR ICIGNVCR SDLLPTKDK THQYISQR IDVPLQIEK INYENALLAR YEVDNNMAQK AKTHQYISQR VYSYYNLDEK VNDDYLIWGSR ACETNVDYVYK HFEVGFIQPGSVK YRINYENALLAR CAGETCSSLNHQER ATMTILTFYNAQLQEK	3FRP_B	31	289
14	Chain B of Cobra Venom Factor	QGTDENPR ICIGNVCR THQYISQR	3FRP_B	31	261

Table 5.4 (continued)

		KCQEALNLK INYENALLAR YEVDNNMAQK VYSYYNLDEK VNDDYLIWGSR ACETNVDYVYK HFEVGFIQPGSVK FYHPDKGTGLLNK CAGETCSSLNHQER WPHEDECQEEEFQK			
15	Chain B of Cobra Venom Factor	QGTDENPR ICIGNVCR THQYISQR IDVPLQIEK INYENALLAR VYSYYNLDEK VNDDYLIWGSR ACETNVDYVYK HFEVGFIQPGSVK FYHPDKGTGLLNK CAGETCSSLNHQER WPHEDECQEEEFQK	3FRP_B	31	240
16	Hemorrhagic metalloproteinase-disintegrin (kaouthiagin)	TNTPEQDRYLQAEK YIEFYVIVDNR IRVYEMINAVNTK VYEMINAVNTK RTAPAFQFSSCSIR TAPAFQFSSCSIR CPTLTNQCIALLGPHFT VSPK GCFDLNMR GCFDLNMRGDDGSFCR GDDGSFCR LYCTEK NTMSCLIPPNPDGIMAE PGTK GQCVDVQTAY	P82942	28	324
17	Snake venom serine protease	IKLGMHSK LKPAVYTK FFCLNTK DIMLIRLR KFFCLNTK TLCAGILQGGK	Q9PT41.1	32	53
18	No significant hit				
19	No significant hit				
20	Chain G, Cobra Venom Factor	VPDTEIETK LKVVPEGVQK SIVTIVKLDPR GICVAEPYEIR DLTEEPNSQGISSK LDDRVPDTEIETK AILHNYVNEDIYVR	3FRP_G	32	48

Table 5.4 (continued)

21	Chain G, Cobra Venom Factor	TMSFYLR VPDTEIETK LKVVPEGVQK GVGGTQLEVIK SIVTIVKLDPR GICVAEPYEIR ASVQEALWSDGVR DLTEEPNSQGISSK LDDRVPDTEIETK AILHNYVNEDIYVR	3FRP_G	32	79
22	Chain A of Natrin, from *Naja atra* CRISP	NVDFNSESTR EIVDLHNSLR RVSPTASNMLK SNCPASCFCR QSSCQDDWIK NVDFNSESTRR EIVDLHNSLRR QKEIVDLHNSLR VLEGIQCGESIYMSSNAR	1XTA_A	25	436
23	Phospholipase A_2, neutral A	CWPYFK GGSGTPVDDLDR NMIQCTVPNR ISRCWPYFK SWWHFADYGCYCGR CCQIHDNCYNEAEKISR LAAICFAGAPYNDNNYNI DLKAR GGSGTPVDDLDRCCQIH DNCYNEAEK	Q92084	45	230
24	Phospholipase A_2, neutral A	CWPYFK GGSGTPVDDLDR NMIQCTVPNR GGNNACAAAVCDCDR TYSYECSQGTLTCK CCQIHDNCYNEAEK LAAICFAGAPYNDNNYNI DLK GGNNACAAAVCDCDRLA AICFAGAPYNDNNYNIDL K	Q92084	52	320
25	Phospholipase A_2, neutral A	CWPYFK GGSGTPVDDLDR NMIQCTVPNR ISRCWPYFK GGNNACAAAVCDCDR TYSYECSQGTLTCK SWWHFADYGCYCGR CCQIHDNCYNEAEKISR LAAICFAGAPYNDNNYNI DLKAR GGSGTPVDDLDRCCQIH	Q92084	43	287

Table 5.4 (continued)

		DNCYNEAEK GGNNACAAAVCDCDRLA AICFAGAPYNDNNYNIDL K			
26	Phospholipase A_2, neutral B	SWWHFADYGCYCGR CCQIHDNCYNEAEK TYSYECSQGTLTCK GGNNACAAAVCDCDR	Q92085	44	395
27	Phospholipase A_2, acidic D	NMIQCTVPNR SWWDFADYGCYCGR GGSGTPVDDLDR GGSGTPVDDLDRCCQVHD NCYGEAEK CCQVHDNCYGEAEK TYSYECSQGTLTCK TYSYECSQGTLTCKGGND ACAAAVCDCDR GGNDACAAAVCDCDR GGNDACAAAVCDCDRLA AICFAGAPYNDNNYNIDL K LAAICFAGAPYNDNNYNI DLK	Q9I900	43	467
28	Phospholipase A_2, acidic D	NMIQCTVPNR SWWDFADYGCYCGR GGSGTPVDDLDR CCQVHDNCYGEAEK CWPYFK TYSYECSQGTLTCK GGNDACAAAVCDCDR LAAICFAGAPYNDNNYNI DLK	Q9I900	42	469
29	Venom nerve growth factor	NPNPEPSGCR IDTACVCVITK ALTMEGNQASWR CKNPNPEPSGCR FIRIDTACVCVITK GIDSSHWNSYCTETDTFIK EDHPVHNLGEHSVCDSV SAWVTK	P01140	26	118
30	Long neurotoxin 7	TWCDGFCSSR	O42257.1	39	67
31	Cardiotoxin from Naja atra	GCIDVCPK YVCCNTDR RGCIDVCPK YVCCNTDRCN MFMVSNLTVPVKR	1CRE_A	45	188

Protein scores greater than 40 are significant ($p < 0.05$)
The mass spectra acquired were searched against all non-redundant NCBI database with taxonomy set to Serpentes (taxid: 8570)

5.4 DISCUSSION

Resource® S ion exchange chromatography resolved the *N. sumatrana* venom into 11 fractions. Further separation of each fraction by reverse-phase HPLC yielded more protein fractions. Fraction 1, the acidic protein fraction yielded poorly resolved multiple fractions (17) in the reverse-phase HPLC. This fraction did not exhibit lethal toxicity and little phospholipase A_2 activity, and presumably consists mainly of enzymes and miscellaneous non-toxic venom proteins.

For fractions 2 to 11, repeated chromatography using reverse-phase HPLC yielded 17 identified protein toxins and more than 20 very minor fractions that could not be identified. The 17 toxins included 8 phospholipases A_2 and 9 three-finger toxins (4 neurotoxins and 5 cardiotoxins).

The seven most abundant proteins are: *N. sumatrana* neurotoxin 1 (a long neurotoxin); *N. sumatrana* neurotoxin 2 (a short neurotoxin), *N. sumatrana* PLA_2-1 (an acidic PLA_2), *N. sumatrana* PLA_2-2 (a neutral PLA_2), and *N. sumatrana* CTX-1, 2 and 3. The seven toxins together account for 87% of total venom protein and they are primarily responsible for the toxic action of the *N. sumatrana* venom. Though the relative abundance of the long neurotoxin and short neurotoxins were only approximately 8.7% and 3.5% respectively, their high lethality (LD_{50} of 0.08 and 0.07 μg/g, respectively) indicates the main toxic action of the venom is presumably still neurotoxic, even though the large amount of cardiotoxins and phospholipases A_2 (relative abundance of 45% and 30%, respectively, with LD_{50} of about 1.0 μg/g and 2.0 μg/g, respectively) certainly contribute significantly to the overall pathophysiological action of *N. sumatrana* venom.

Thus, the Resource® S coupled to reverse-phase HPLC approach not only identified the major/minor toxins of the cobra venom but also yielded semi-quantitative data on the relative abundance of the toxins. These information are important in the understanding of the pathophysiological action of the venom. Besides, the method is also semi-preparative yielding small amount of reasonably purified toxins (homogeneous as indicated by SDS-PAGE).

Shotgun LC-MS/MS coupled with *de novo* sequencing revealed the presence of 50 individual proteins in *N. sumatrana* venom, with three-finger toxins (both neurotoxins and cardiotoxins) and phospholipase A_2 constitute about 38% and 36%, respectively of the types of proteins identified. In contrast, Li *et al*. (2004) reported the identification of 78 proteins in *Naja atra* venom using shotgun LC-MS/MS, in which three-finger toxins and phospholipase A_2 constitute about 61% and 18%, respectively, and the most numerous type of venom proteins detected were cardiotoxins (44), accounting for 56% of total proteins identified. However, in the present studies, phospholipase A_2 which account for 36% (or a total of 18 proteins) of total proteins identified were the most numerous type of toxins while cardiotoxins (9 in total) only account for 18% of the type of venom proteins identified. Besides these two families of low molecular weight venom toxins, Li *et al*. (2004) also reported the presence of nerve growth factor, cobra venom factor, kaouthiagin, natrin (CRISP protein), serum albumin, hemorrhagic metalloproteinase, Ohanin and Thaicobrin. These proteins were also identified in shotgun LC-MS/MS of *N. sumatrana* venom, except for kaouthiagin (which was identified in *N. sumatrana* venom in the latter gel filtration-2DE approach) and Ohanin. Even though Ohanin was not detected in *N. sumatrana* venom, the highly homologous thaicobrin was detected. The present shotgun analysis revealed the presence of aminopeptidase, complement depleting factors,

natriuretic peptide and NADH-dehydrogenase that were not reported by Li *et al*. (2004) in *N. atra* venom. Kulkeaw *et al.* (2007) reported that 2DE-LC-MS/MS shotgun analysis of *Naja kaouthia* venom revealed the presence of 61 proteins in the venom, again, with three-finger toxins and phospholipase A_2 constitute 48% and 31%, respectively, of the identified proteins. Phospholipases A_2 were the most numerous type of proteins (31%) and there were only 12 cardiotoxins detected, accounting for 20% of total types of proteins identified. This is similar to protein distribution for *N. sumatrana* venom. Kulkeaw *et al*. (2007) also reported the presence of cobra venom factor, CRISP, serum albumin and kaouthiagin.

All three shotgun LC-MS/MS of the Asiatic cobra venoms revealed the presence of large number of three-finger toxins, phospholipase A_2 and the presence of CRISP, cobra venom factor, serum albumin and hemorrhagic metalloproteinase. These proteins appear to be the common components of Asiatic cobra venoms. It should be noted, however, although shotgun LC-MS/MS allows global protein identification of cobra venom proteins, it generally does not give accurate information regarding the relative abundance of the various toxins and proteins, the knowledge of which is essential for the understanding of the pathophysiological action of snake venom.

The above mentioned multidimensional chromatographic approaches did not identify any high molecular weight protein components (mostly enzymes and cobra venom factors) of the venom, many of which were detected by the shotgun analysis. This suggests that the high molecular weight proteins were possibly in the acidic fraction (Fraction 1) of the Resource® S ion exchange chromatogram, which accounts for 4% of total venom proteins. Reverse-phase HPLC of the fraction yielded multiple yet poorly resolved peaks. Since it is

known that the high molecular weight venom proteins exist mostly in minute amount, identification of these proteins by chromatographic fractionation is likely to be difficult. As such, a different approach was adopted in which Sephadex® G-50 gel filtration chromatography was used to isolate the high molecular weight protein fractions (which consist mostly of non-toxic enzymes and proteins), and the isolated high molecular weight fraction was then subjected to 2DE with silver staining to visualize the protein spots. MALDI TOF/TOF analysis of the spots revealed the presence of 9 low molecular weight proteins (MW <20 kDa), (all except one have been identified by the multidimensional chromatography approach), as well as other known cobra venom proteins including cobra venom factor, complement depleting factor, serum albumin precursor, phosphodiesterase, 5'-nucleotidase, hemorrhagic metalloproteinase, serine proteinase, natrin and venom nerve growth factor were also presence. It is noteworthy that of the 31 spots examined, 7 spots were identified either as cobra venom factor or subunit of the protein. Since the cobra venom factor, cobra venom factor chain B and chain G all appear as train of spots, it is possible they are the same protein with different post-translational modification (Fu and Gowda, 2001; Gowda *et al*., 2001). From the data available, it is not possible to conclude that there is only one form of cobra venom factor with different degree of glycosylation or there are multiple forms of the protein. However, the data suggest the presence of numerous cobra venom factors in the venom. Two unusual proteins identified were oligophrenin-1 and plastin-3 that exhibited conserved sequence homologs with the transcripts of *Crotalus adamanteus* venom gland (Rokyta *et al*., 2012) but not present in other cobra venoms. The detection of the various enzymes (phosphodiesterase, 5'-nucleotidase, serine protease, hemorrhagic metalloproteinase) and other known

cobra venom proteins (nerve growth factor, serum albumin, complement-depleting factor) indicated these proteins are presence in detectable though minute amount in the cobra venom.

Overall view of the proteome of *N. sumatrana* venom

To date, among the Asiatic cobra, only the proteome of venoms from two non-spitting cobras, *Naja naja atra* (Chinese cobra) and *Naja kaouthia* (Thai cobra) were investigated. Li *et al*. (2004) reported a combination of four different approaches *i.e.* shotgun LC-MS/MS, 1DE-LC-MS/MS, Gel filtration-LC-MS/MS and Gel filtration-2DE-MALDI-TOF-MS to investigate the proteome of Chinese cobra. They reported identification of 124 proteins and peptides in the cobra venom. Kulkeaw *et al*. (2007), on the other hand, used 2DE coupled to liquid chromatography and tandem mass spectrometry to investigate the proteome of the Thai cobra and identified 61 proteins in the venom.

In the present work, two approaches were used: shotgun analysis (enhanced by *de novo* sequencing), which revealed the presence of 50 proteins in the venom, and a combination of multi-dimensional chromatography and gel filtration-2DE (abbreviated as multi-D chromatography-2DE) approaches, which together revealed the presence of 30 venom proteins. The shotgun analysis was able to detect the presence of very minor protein components, whereas the multi-D chromatography-2DE detected primarily the more abundance venom proteins, with semi-quantitative information. Because of this, the latter approach could provide a better understanding of the pathophysiological action of the venom. Comparison of the accession number of the proteins identified suggested that there are only 5 proteins detected by both approaches: the cobra venom factor (Q91132), CTX-2b (Q9PST3), CTX-4b (O73856), the complement depleting

factor (ABN72543) and the acidic phospholipase A_2-D (Q9I900). However, since the number of peptides detected by shotgun analysis was relatively limited, the number of overlapping proteins could well be more than that. The multidimensional chromatographic approach showed that the phospholipases A_2 and three-finger toxins together account for 90% of the total venom protein by weight, in terms of relative abundance.

The remaining 10% of the venom protein consists of venom enzymes and non-enzymatic proteins. The enzymes detected in *N. sumatrana* venom include cobrin, kaouthiagin, serine protease, aminopeptidase, phosphodiesterase, 5'nucleotidase and NADH-dehydrogenase. However, other common venom enzymes such as acetylcholinesterase, L-amino acid oxidase, alkaline phosphomonoesterase and hyaluronidase were not detected despite the earlier biochemical characterization of *N. sumatrana* venom showed the presence of these enzymes (Chapter 4). These enzymes were not detected by the proteomic approach presumably because of the limited mass spectrometry database on venom enzymes. Phosphodiesterase may cause reduction of mean arterial pressure and locomotor depression, presumably due to the reduction of cAMP levels (Russell *et al*., 1963). Enzyme 5'-Nucleotidase could inhibit platelet aggregation and may contribute to the disturbance of hemostatic functions (Ouyang and Huang, 1983 and 1986). The presence of hemorrhagic metalloproteinase-disintegrin such as kaouthiagin has also been reported in *N. kaouthia* venom (Ito *et al*., 2001). This enzyme probably belongs to the SVMP family of enzyme, an important family of enzyme in crotalid venom but only presence in very small amount in cobra venom (Petras *et* al., 2011). Another metalloproteinase-disintegrin detected was cobrin, which can cleave complement protein C3 to C3C-like (O'Keefe *et al*., 1988). Both kaouthiagin

and cobrin exist in very small amount and hence probably not important in the pathophysiological action of the venom.

Besides enzymes, there are also several cobra venom proteins detected in *N. sumatrana* venom that appear to be common components of Asiatic cobra venoms as these were also detected in *N. atra* and *N. kaouthia* venoms. These include Thaicobrin, kaouthin-1, natrin (a CRISP), natriuretic peptide, complement depleting factors, cobra venom factors, venom nerve growth factor and serum albumin. Some of these proteins are known to exhibit potent pharmacological action. Thaicobrin belongs to vespryn family that is similar to Ohanin from *Ophiophagus hannah* (Pung *et al.*, 2005), a protein that produces dose-dependent hypolocomotion and hyperalgesia in mice. Kaouthin-1 exhibits ion channel blocking activity (Matsunaga *et al.*, 2009). Natrin, a member of the CRISP family and has been reported as a BKCa channel blocker (Wang *et al*., 2006). Natriuretic peptide is involved in natriuresis, diuresis and vasodilation (Levin *et al*., 1998). Nerve growth factor (NGF) from snake venom shares 64% sequence homology with mammalian's NGF that is important to stimulate differentiation and maintenance of the sympathetic and embryonic sensory neurons. However its role in the snake venom is not clear and probably involves directly or indirectly in the toxic action of snake venom (Kostiza and Meier, 1996).

Cobra serum albumin appears to exhibit anti-toxic properties (Shao *et al*., 1993 and Wang *et al*., 1998). These proteins exist also in minute amount in cobra venom and their roles in the pathophysiological action of the venom are likely to be limited.

On the other hand, gel filtration-2DE data suggest the presence of appreciable amount of cobra venom factors in the high molecular weight fraction of the

venom. Cobra venom factors and complement depleting factors are both involved in activation of complement proteins, which is responsible to mediate the local inflammatory processes, indicating that they may play an important role in causing the local inflammation and tissue damages in cobra bites (Vogel and Fritzinger, 2010; Warrell *et al*., 1976). Two-DE of high molecular weight venom protein fractions, however, revealed the presence of two uncommon proteins: oligophrenin-1 and plastin-3. These two proteins have not been reported in other cobra venom. It was, however also present in *Crotalus adamanteus* venom (Rokyta *et al*., 2012). Oligophrenin-1 is probably related to Rho-GTPase-activating protein, involved in intracellular signal transductions. Plastin-3 belongs to a family of actin-binding proteins that are conserved throughout eukaryote evolution (NCBI Gene ID: 5358, updated on 28 April 2013), however its role in toxic action of venoms remain unclear.

Comparison of proteome of Equatorial spitting cobra and African spitting cobra venoms

Petras *et al.* (2011) investigated the proteome of venoms from 6 African spitting cobras using reverse-phase HPLC separation followed by characterization of the venom fractions by N-terminal sequencing, ESI mass spectrometry and SDS-PAGE. They identified 22-32 proteins in the various venoms. For example, for *N. nigricollis*, 32 proteins were identified, including 21 cardiotoxins, 4 neurotoxins and just one type of phospholipase A_2. In terms of abundance, the three-finger toxins (mainly cardiotoxins) account for about 70% of the venom protein, and phospholipases A_2, 22-30%. The amount of neurotoxins was between trace to 4.4%, except for *N. nubiae* that has a neurotoxin content of 12.6%. Together the three-finger toxins and phospholipases A_2 account for 95-

98% of the total HPLC-separated proteins. The remaining proteins identified include PIII-SVMP, Nawaprin (only in certain *N. nigricollis*), CRISP and endonuclease. In contrast, for *N. sumatrana* venom, the relative abundance of cardiotoxin is much lower (45% of venom protein, compared to 60-70% in African spitting cobra venoms), with comparable amount of phospholipases A_2 and a higher neurotoxins content (comparable to venom of *N. nubiae*). In term of type of proteins, *N. sumatrana* contains far greater types of phospholipases A_2 and neurotoxins but far less types of cardiotoxins compared to African spitting cobra venom. A similar toxin composition was also reported in *N. kaouthia* venom: there are 19 phospholipases A_2 but only 12 cardiotoxins (Kulkeaw *et al*., 2007). On the other hand, in *N. atra* venom, there are 14 phospholipases A_2 but 44 cardiotoxins, out of the 78 venom proteins identified (Li *et al*., 2004). The relatively small number of phospholipases A_2 detected for the African spitting cobra venom could partly due to the limitation of the sensitivity of the separation method used. Nevertheless, this comparison does show that Asiatic cobra venom appears to have much larger number of phospholipases A_2. The African spitting cobra venoms usually contain very large amount of cardiotoxins but less than 5% of neurotoxin, whereas for *N. sumatrana* venom, while the cardiotoxin is still the most abundant venom protein, its content is lower (45%) and neurotoxin content is higher. These differences in toxin distribution suggest possible differences in the pathophysiological action of *N. sumatrana* venom and the African spitting cobra venoms.

Many minor, high molecular weight venom proteins detected in this work were not reported in African spitting cobra venom, this could either be due to the differences in the sensitivity of the methods of separation and detection used, is reflective of the true difference between Asiatic cobra venoms and African

cobra venoms. It is interesting to note that the cobra venom factor is absent in African spitting cobra venoms. More detailed studies of more Asiatic and African cobra venom are required to resolve this issue. It is interesting to note that despite the relatively large number of venom toxins found in the African spitting cobra venoms, in each case, only 6-8 toxins have relative abundances of more than 5% of the venom proteomes. The same situation was also observed in *N. sumatrana* venom, where only 6 toxins have relative abundances of more than 3%.