Characterization of bioactive peptides without disulfide bridges from the venom of *Lycosa poonaensis* species inhabiting the Egyptian environment

Alhussin Mohamed Abdelhakeem Megaly

2022

CONTENTS Page

CHAPTER 1

General introduction

Spider

Spiders (Arachnida: Araneae) are one of the most successful group of organisms on Earth (King & Hardy, 2013), which first evolved around 300 million years ago. Currently, more than 50,000 species are distributed worldwide according to the World Spider Catalog (June 2022, https://wsc.nmbe.ch/) (Langenegger, Nentwig, & Kuhn-Nentwig, 2019; Vassilevski, Kozlov, & Grishin, 2009). Among them, only about 0.5% of all species are considered to be dangerous to humans (Windley et al., 2012). Spiders differ from insects in having eight legs instead of six, and the body is divided into two parts and possess silk gland (Figure 1-1). The silk web is built to capture or wrap the prey to stop their movement (Kuhn-Nentwig, Stocklin, & Nentwig, 2011). In addition to the web, spiders use venom for predation and defencing. These characteristics have allowed spiders to flourish for such a long period. The venom glands are known to be present in most spider species, except for a few spider families. The venom gland is located in the basal segment of the chelicera of mygalomorph spiders or in the prosoma of araneomorph spiders (Kuhn-Nentwig, et al., 2011; Langenegger, et al., 2019).

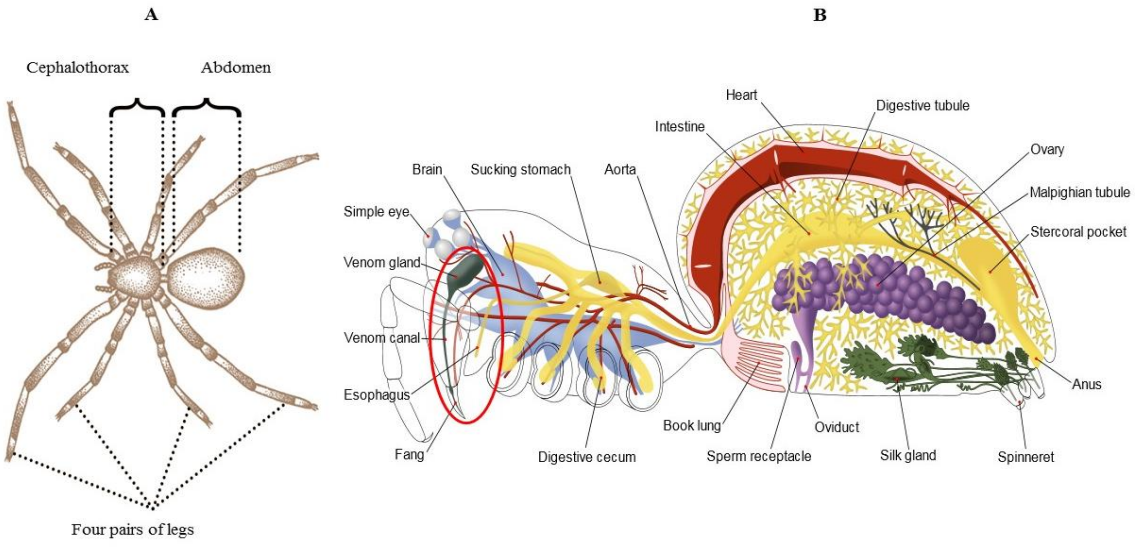


Figure 1-1. A) Spider image from CDC Public Health Image Library (public domain) shows the main characteristics of Arachnids (General body parts), B) (Edited photo) the original diagram from the spider book (1912, 1920) by John Henry Comstock, (anatomical information of spider body), the red circle indicate the position of the venom gland and venom canal.

Egyptian Spiders

The history of the scientific record of a spider species from Egypt was discussed by El-Hennawy (El-Hennawy, 2000), in which he mentioned that the first record was that of Linnaeus in his 10th edition of "Systema Naturae". He described two spider species, but the description was very short and not enough to identify such species as Simon stated. The second record was done by Forskål, who described four spider species from the region of Cairo with sufficient description to identify these species. The third and the real scientific study of Egyptian Arachnida start with the historical work "Description de l'Égypte", where we can find the first landmark in the study of Arachnida of Egypt. It includes the work of Savigny on Egyptian arachnids, which was completed and edited by Audouin. The story of this work was summarized by Simon. He discussed and highlighted the two authors' life and contribution in the scientific record of arachnids.

El-Hennawy published "Annotated checklist of Egyptian spider species" in 1990 and followed it by "A list of Egyptian spiders (revised in 2002)" and finally to bring these data to be up to date he published "A list of Egyptian spiders (revised in 2006)" which showed the record of 40 spider families belonging to 193 genera and 385 species (Figure 1-2) and their locality (Figure 1-3) (El-Hennawy, 2006).The Egyptian spiders are still incompletely known due to scarcity of studies on this group. In addition, most of the recorded species were identified based on the morphological characters, which raise a high possibility for misidentification of many recorded species.

Order Araneida (Araneae, Aranei) Suborder Opisthothelae Infraorder Mygalomorphae					
Nemesiidae	1 (1)	Theraphosidae	1 (3)		
Infraorder Araneomorphae					
Agelenidae	5 (7)	Liocranidae	1 (2)	Scytodidae	1 (5)
Araneidae	15 (22)	Lycosidae	20 (44)	Segestriidae	2 (2)
Cithaeronidae	1 (1)	Mimetidae	1 (1)	Selenopidae	1 (1)
Clubionidae	1 (1)	Miturgidae	2 (9)	Sicariidae	1 (1)
Corinnidae	1 (1)	Oecobiidae	2 (7)	Sparassidae	6 (13)
Ctenidae	1 (1)	Oonopidae	4 (5)	Synaphridae	1 (1)
Dictynidae	5 (6)	Oxyopidae	2 (6)	Tetragnathidae	2 (5)
Dysderidae	1 (7)	Palpimanidae	1 (3)	Theridiidae	10 (24)
Eresidae	3 (6)	Philodromidae	3 (18)	Thomisidae	10 (25)
Filistatidae	2 (3)	Pholcidae	5 (5)	Titanoecidae	2 (2)
Gnaphosidae	21 (48)	Pisauridae	4 (4)	Uloboridae	1 (2)
Hersiliidae	2 (2)	Prodidomidae	3 (3)	Zodariidae	5 (8)
Linyphiidae	8 (8)	Salticidae	35 (72)		

Figure 1-2. The recorded spiders family in Egypt. The number of genera and species were shown.

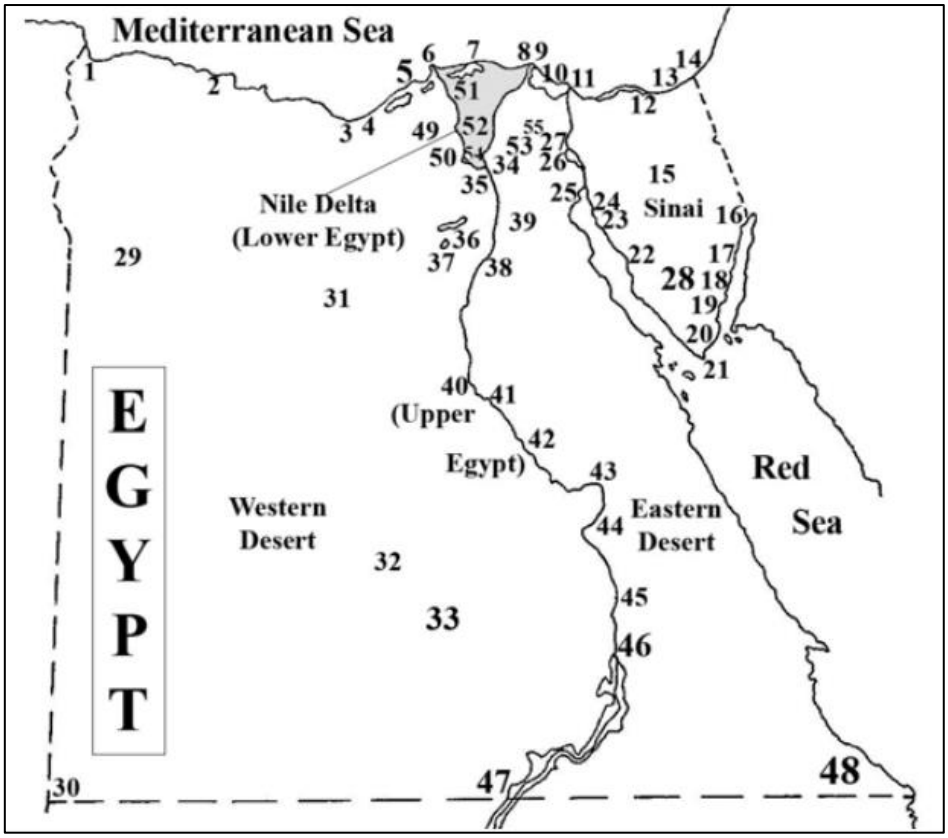


Figure 1-3. Distribution of spider in Egypt. 1- El-Sallum. 2- Marsa Matruh. 3- El-Omayed Prot. 4- ElHammam. 5- Alexandria, Edko, Mariout. 6- Rosetta. 7- El-Burullus Prot. 8- Ras El-Barr. 9- Damietta. 10- El-Manzalah (lake). 11- Port Said. 12- El-Zaranik Prot. 13- ElArish. 14- Rafah. 15- Mid Sinai. 16- Taba. 17- Abu Galoum Prot. 18- Dahab and Wadi Yah'med. 19- Nabq Prot. 20- Sharm ElSheikh. 21- Ras Mohammed Prot. 22- En Higiya (north east of Abu Zneima). 23- Ras Sedr. 24- AinMusa. 25- Suez. 26- Fayed. 27- Ismailia. 28- St. Catherine, Mount Serbal, Wadi Esla. 29- Siwa Oases. 30- El-Uwaynat. 31- El-Baharia Oases, ElBawitti. 32- Dakhla Oases. 33- New Valley. 34- Cairo (Heliopolis, Zenhum and Helwan), Wadi Degla (El-Maadi). 35- Giza, Pyramids, Saqqarah, Dahshur. 36- El-Fayum, Kom Osheem. 37- Wadi El-Raiyan. 38- Beni Suef. 39- Wadi Rishrash. 40- Manfalut. 41- Assiut. 42- Sohag. 43- Qena. 44- Luxor. 45- Gebel Silsilis. 46- Aswan, Elephantine and Philoe island; Fatira (Kom Ombo). 47- Wadi-Halfa, Nubia. 48- El-Shalateen, Bir El-Gahliya, Wadi De'eeb. 49- Wadi Natron. 50- El-Tahrir Province. 51- Kafr El-Sheikh. 52- El-Menoufeia, Shebin El-Kom. 53- El-Aasher-Min-Ramadan City (65 km east of Cairo). 54- Nile Barrage, Qalyubia. 55- Salahyeh.

Spider venom

Spider venom is a complex mixture of bioactive components, in which peptides play an important role by displaying various functions, such as neurotoxic, analgesic, cytotoxic, necrotic activities (Windley, et al., 2012). Components of spider venom are typically divided into four groups; (1) small molecular mass compounds (SMMCs), (2) antimicrobial peptides (AMPs), (3) peptide neurotoxins, and (4) proteins and enzymes (Langenegger, et al., 2019).

Among them, disulfide-rich peptides have evolved as a neurotoxin to effectively act on specific ion channels or receptors in prey animals. Spider venoms typically contain dozens of different cysteine-rich peptides and typically have molecular masses between 3 and 9 kDa, and feature ≥6 cysteine residues in conserved structural motifs (Kuhn-Nentwig, et al., 2011; Langenegger, et al., 2019)

AMPs also have critical functions in the venom, although they have been identified from limited spider families. They are mainly responsible for disrupting the integrity of the cellular membrane of bacteria and preventing the infection of pathogens during injection of venom into prey or ingestion of prey (Windley, et al., 2012). In addition, some AMPs exhibit insecticidal activity, but also enhance the action of other neurotoxins.

SMMCs are considered to be present in most spider venoms, although they are only marginally investigated compared to venom peptides (Langenegger, et al., 2019). They include inorganic ions, organic acids, nucleotides, nucleosides, amino acids, amines, and polyamines. Some SMMCs play a role in synergistically enhancing the activity of venom peptides. The high potassium content of spider venom can cause depolarization of excitable cell membranes, leading to paralysis of the prey. Also, existence of organic acids, primarily citric acid, which is possibly the most common organic acid present in spider venom, makes spider venom acidic with pH values between 5.3 and 6.1. In addition, citric acid was proposed to reversibly inhibit divalent cation dependent enzymes in the venom gland by complexation of Ca^{2+} and Zn^{2+}. After venom injection, citric acid is diluted to shift the equilibrium in direction of complexation of the cations by the venom enzymes, leading to reactivation of the enzymes.

Animal venom study

The human population is expected to reach about 9.3 billion in 2050 and 10.1 billion in 2100. A stable and affordable food supply is critical to accommodate this population growth. However, arthropod pests, primarily insects, reduce the world crop yields by 10-14% despite the advent of chemical insecticides to combat insect pests. Although chemical insecticides remain the main approach for controlling insect pests, the resistance to the insecticides have been evolved after prolonged use. The development of pest resistance can be reduced by rotation of different insecticides. However, the development of new insecticides takes much longer time than before, limiting the choice of insecticides that can be used. In this context, insecticidal peptides derived from animal venom are attracting attention as biopesticides (Koua et al., 2020).

These peptides typically have the advantage of high selectivity for target pests and low environmental toxicity, which leads to no adverse effects on beneficial insects such as honeybees. On the other hand, with a growing population, the spread of new infectious diseases becomes a significant problem for humans. The emergence of antibiotic-resistant pathogens and the decreased rate of development of new antibiotics are a serious threat to human health. AMPs from animal venom could be an alternative to conventional antibiotics (Irazazabal et al., 2016). Thus, these increasing interests in toxins from animal venom resulted in several studies being considered a source of new agrochemicals and pharmaceuticals (Koua, et al., 2020; Windley, et al., 2012).

Aim and objectives of this study

The purpose of this study is to gain a better understanding of the bioactive components in the venom of the spider *Lycosa poonaensis* found in Egypt.
To achieve this objective, the study focused on the following points.

1. Characterization of linear peptides of the *L. poonaensis* venom by mass spectrometric analysis (Chapter 2).
2. Comprehensive analysis of components of the *L. poonaensis* venom by transcriptome analysis (Chapter 3).
3. Evaluation of biological activities of linear peptides identified by transcriptome analysis of the *L. poonaensis* venom gland (Chapter 4).

The results obtained in this study will not only contribute to our deeper understanding of the function of linear peptides in spider venom, but will also provide useful insights in the development of novel biopesticides and pharmaceuticals.

References

El-Hennawy, H. K. (2000). The first landmark in the route of Egyptian Arachnology:" Explication Sommaire des Planches d'Arachnides de l'Égypte et de la Syrie"(1825). *Serket, 6*, 115-128.

El-Hennawy, H. K. (2006). A list of Egyptian spiders (revised in 2006). *Serket, 10*(2), 65-76.

Irazazabal, L. N., Porto, W. F., Ribeiro, S. M., Casale, S., Humblot, V., Ladram, A., & Franco, O. L. (2016). Selective amino acid substitution reduces cytotoxicity of the antimicrobial peptide mastoparan. *Biochimica Et Biophysica Acta-Biomembranes, 1858*(11), 2699-2708. doi: 10.1016/j.bbamem.2016.07.001

King, G. F., & Hardy, M. C. (2013). Spider-Venom Peptides: Structure, Pharmacology, and Potential for Control of Insect Pests. *Annual Review of Entomology, Vol 58, 58*, 475-496. doi: 10.1146/annurev-ento-120811-153650

Koua, D., Mary, R., Ebou, A., Barrachina, C., El Koulali, K., Cazals, G., . . . Dutertre, S. (2020). Proteotranscriptomic Insights into the Venom Composition of the Wolf SpiderLycosa tarantula. *Toxins, 12*(8). doi: ARTN 50110.3390/toxins12080501

Kuhn-Nentwig, L., Stocklin, R., & Nentwig, W. (2011). Venom Composition and Strategies in Spiders: Is Everything Possible? *Advances in Insect Physiology, Vol 40: Spider Physiology and Behaviour - Physiology, 40*, 1-86. doi: 10.1016/B978-0-12-387668-3.00001-5

Langenegger, N., Nentwig, W., & Kuhn-Nentwig, L. (2019). Spider Venom: Components, Modes of Action, and Novel Strategies in Transcriptomic and Proteomic Analyses. *Toxins, 11*(10). doi: ARTN 61110.3390/toxins11100611

Vassilevski, A. A., Kozlov, S. A., & Grishin, E. V. (2009). Molecular diversity of spider venom. *Biochemistry-Moscow, 74*(13), 1505-1534. doi: 10.1134/S0006297909130069

Windley, M. J., Herzig, V., Dziemborowicz, S. A., Hardy, M. C., King, G. F., & Nicholson, G. M. (2012). Spider-Venom Peptides as Bioinsecticides. *Toxins, 4*(3), 191-227. doi: 10.3390/toxins4030191

CHAPTER 2

Characterization of two linear peptides without disulfide bridges from the venom of the spider *Lycosa poonaensis* (Lycosidae)

2.1 Introduction

Spiders are an ancient group of arthropods that evolved more than 400 million years ago (Lozano-Fernandez et al., 2016). The evolutionary success of spiders can be attributed partly to their venom, which has been used to capture prey and defend themselves against predators. Spider venom is a complex mixture of bioactive components, which show multiple functions, such as neurotoxic, analgesic, cytotoxic, and necrotic activities (Akef, 2018; Kuhn-Nentwig et al., 2011; Langenegger et al., 2019). Thus, there is growing interest in spider venom as a potential source of novel insecticides and drugs (Windley et al., 2012; Wu et al., 2019). Components of spider venom can be classically divided into several groups according to their chemical nature: (1) low molecular mass compounds, (2) linear peptides without disulfide bridges, (3) cysteine-rich peptides, and (4) proteins and enzymes. These venom components can disrupt transport and signal transduction through the biological membranes involved in various physiological processes. Many linear peptides without disulfide bridges have been identified as antimicrobial peptides (AMPs) (Santos Daniel M. et al., 2016).

Most AMPs in spider venom are cationic and adopt an amphipathic α-helical structure with molecular sizes ranging from 10 to 80 amino acid residues. AMPs disrupt the integrity of the cellular membrane, in which their cationic nature plays a vital role in its interaction with the anionic surface of bacterial membranes to form pores (Avci et al., 2018). From a functional standpoint, AMPs are thought to prevent the infection of pathogens during injection of venom into prey or ingestion of prey (Wullschleger et al., 2005; Yans and Adams, 1998).

Various AMPs have been identified from the venom of the Lycosidae spiders, such as *Hogna carolinensis* (Yans and Adams, 1998), *Lycosa erythrognatha* (Santos D. M. et al., 2010), *L. singoriensis* (Budnik et al., 2004), and *L. sinesis* (Tang et al., 2020). This suggests that AMPs have some critical functions in the venom of the Lycosidae spiders. In this study, we characterized non-disulfide-bridged peptides in the venom of the spider *L. poonaensis* that have never been studied before, using mass spectrometry to find novel AMPs.

2.2 Materials and methods

2.2.1 Collection of spider venom

The *L. poonaensis* spiders were collected in the Western Desert of Egypt. The spider was identified based on the morphological characteristics and the DNA sequence of the mitochondrial cytochrome oxidase subunit I gene (Ashfaq et al., 2019). The venom glands were dissected and manually homogenized in distilled water for 10 min at room temperature. The homogenate was centrifuged at 14,000 × g at 4°C, and the supernatant was lyophilized to obtain the crude venom.

2.2.2 Mass spectrometric analysis

LC/MS^n measurements were carried out on an LCMS-IT-TOF mass spectrometer (Shimadzu, Kyoto, Japan) equipped with an electrospray ion source. A reversed-phase HPLC (RP-HPLC) column (TSKgel ODS 150 x 1.0 mm, Tosoh, Tokyo, Japan) was used for separation. The column was eluted with 0.1% formic acid in H_2O (solvent A) and 0.1% formic acid in CH_3CN (solvent B) at a flow rate of 0.05 ml/min using a linear gradient of 5-60% of solvent B over 55 min.

MALDI-TOF/TOF-MS measurements were performed on an Autoflex III mass spectrometer (Bruker Daltonics, Bremen, Germany) using LIFT mode under high-energy CID conditions with argon as collision gas. Samples were dissolved in a matrix solution containing 10 mg/mL of 2,5-dihydroxybenzoic acid or α-cyano-4-hydroxycinnamic acid in 50% acetonitrile containing 0.1% TFA. The solution was spotted onto the target plate and allowed to dry at room temperature. External calibration of the mass scale was carried out using peptides of known molecular masses. Interpretation of the MS/MS spectra was carried out manually with the aid of an open-source software mMass (Strohalm et al., 2010).

2.2.3 HPLC fractionation

A crude venom sample (50 μg) was injected into an RP-HPLC column (Everest C18 250 x 1.0 mm, Hichrom, Reading, UK). The column was eluted with solvent A and solvent B at a 0.05 ml/min flow rate using a linear gradient of 5-60% of solvent B over 55 min. The elution was monitored by UV absorbance at 215 nm. Fractions were manually collected every 5 min. Fraction eluting at the retention time of 35-40 min was further fractionated using a linear gradient of 5-60% of solvent B over 90 min.

2.2.4 Reduction and alkylation of Cys residues

Crude venom (100 μg) was dissolved in the buffer containing 0.13 M $NaHCO_3$ (pH 8.5), 2.7 M urea, and 35 mM dithiothreitol (DTT). The mixture was incubated for 1 h at 50°C before reacting with iodoacetamide (final concentration, 125 mM) for 1 h at 25°C. LC/MS analyzed the samples without purification.

2.2.5 Peptide Synthesis

Peptides were synthesized by the Fmoc-based solid-phase method on a Rink-amide resin (0.17-0.19 mmol/g). Fmoc deprotection was carried out with 20% piperidine in DMF three times, each for 3, 3, and 20 minutes at room temperature. The reaction was monitored by the Kaiser test (Kaiser et al., 1970). Each Fmoc-protected amino acid (3 eq) were coupled using 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) and *N*,*N*-diisopropylethylamine (DIPEA) in DMF under microwave heating for 5 min at 75°C (Initiator$^+$, Biotage, Uppsala, Sweden). For coupling of histidine, the reaction was done using 1-hydroxybenzotriazole monohydrate (HOBt•H_2O) and *N*,*N*'-diisopropylcarbodiimide (DIPCDI) at 50℃ for 10 min. After completion of the coupling reaction, the resin was washed 5 times each with DMF, DCM, and diethyl ether and dried in vacuo. The cleavage of peptides from the resin and removal of side-chain protecting groups were performed using the solution containing TFA/triisopropylsilane/water (95/2.5/2.5 by volume). After incubation for 30 min on ice followed by 90 min at room temperature, the resin was removed by filtration. Cold diethyl ether was added to the filtrate to precipitate the peptides. The peptide was washed with cold diethyl ether twice and dried in vacuo. HPLC purified the desired peptide on a C18 column. The identity of the peptides was confirmed by MS analyses as described above.

2.2.6 Antibacterial activity

Antibacterial activity was measured using Gram-negative bacteria, *Escherichia coli* NBRC 3972, Gram-positive bacteria, *Staphylococcus aureus* NBRC 13276, *Bacillus subtilis* NBRC 3009 (NITE Biological Resource Center, Chiba, Japan) by a liquid growth inhibition assay. In aerobic conditions, the bacteria were cultivated overnight at 37°C in a petri dish. Each bacteria strain was grown in liquid LB medium (1% tryptone, 0.5% yeast extract, and 1% NaCl). Minimal inhibitory concentrations (MICs) were determined using a 2-fold microtiter broth dilution assay.

Aliquots of each sample (10 µL) were incubated with 90 µL of a suspension of a mid-logarithmic phase culture of bacteria in a 96-well plate at a starting absorbance of A_{595}=0.001 in LB medium overnight at 37°C with continuous shaking. Growth inhibition was monitored by measuring the absorbance at 595 nm using a Benchmark microplate reader (Bio-Rad, Hercules, CA). Experiments were performed in duplicate and repeated at least three times. MICs are expressed as the interval of concentrations, [a]-[b], where [a] is the highest concentration at which bacteria still grow, and [b] the lowest concentration causing growth inhibition.

2.2.7 Insect toxicity

Insect toxicity was tested using crickets (*Acheta domestica*, 50±5 mg body weight) by injection of 1-2 µl of each sample dissolved in distilled water into their abdominal region. Distilled water was injected as a control. For each measurement, 10 insects were used, and the number of paralyzed insects was counted 30 min after injection. The doses required to induce paralysis in half of the test animals (ED_{50}) were determined using a statistical software PRISM (GraphPad Software, La Jolla, CA).

2.2.8 Hemolytic activity

Fresh sheep red blood cells (sRBCs) were washed three times with PBS (35 mM phosphate buffer and 150 mM NaCl, pH 7.2) by centrifugation at 2000 × g for 5 min and resuspended in PBS. Aliquots of each sample in PBS (50 µl) were added to 50 µl of sRBC suspension [4% (v/v) in the final] in a microtube and incubated for 1 h at 37˚C. The samples were centrifuged at 2000 × g for 5 min. The supernatant was transferred to a 96-well plate to monitor hemoglobin release by measuring the absorbance of the supernatant at 450 nm using a Benchmark microplate reader. PBS or 0.1% Triton X-100 were used as the negative and positive controls, respectively. The experiments were performed in triplicate and repeated 3 times. The equation calculated hemolysis rate, where A_P refers to peptide sample absorbance, A_T is 0.1% Triton X-100 positive control absorbance, and A_0 indicates PBS sample negative control absorbance.

$$\text{Hemolysis rate (\%)} = \frac{A_p - A_0}{A_T - A_0} \times 100$$

2.2.9 Circular dichroism (CD) measurements

CD spectra were recorded on a J-820 spectropolarimeter (Jasco, Tokyo, Japan). The spectra were measured between 190-260 nm (0.2 nm step) at 25ºC with a 1-mm-

pathlength cell. Samples were dissolved in 0.2 M phosphate buffer (pH 7.0) or 50% TFE in 0.2 M phosphate buffer (pH 7.0) at a concentration of 10 μM. Five scans were averaged for each sample. The percentages of secondary structure were estimated from the spectra using the DichroWeb server (Whitmore and Wallace, 2008). The helical wheel projection was performed using online software HeliQuest (https://heliquest.ipmc.cnrs.fr).

2.3 Results and Discussion

2.3.1 Components of the L. poonaensis *venom*

The crude venom of the spider *L. poonaensis* was analyzed by LC/MS to gain a general picture of the components. The result revealed that the *L. poonaensis* venom contains over 400 components, whose molecular masses range from 512.3 to 7,381.3 Da (Figure 2-1 and Table 2-1). Next, the number of linear peptides without disulfide bridges was investigated. To specify non-disulfide-bridged peptides, all venom components performed a reduction and alkylation reaction. After this reaction, only disulfide-bridged peptides showed an increase in molecular mass due to Cys alkylation. The components without mass shift were considered a non-disulfide-bridged peptide. By comparing the LC/MS data before and after the reactions, more than 120 components did not show any molecular mass shifts due to Cys alkylation, which were determined to be non-disulfide-bridged peptides (Figure 2-1 and Table 2-1). Molecular masses of the non-disulfide-bridged peptides ranging from 512.3 to 3407.7 Da, and those from 1,000-2,000 Da were the most abundant. This is consistent with previous reports that showed the molecular mass distribution of AMPs in the Lycosidae spider venoms (Kuhn-Nentwig, et al., 2011).

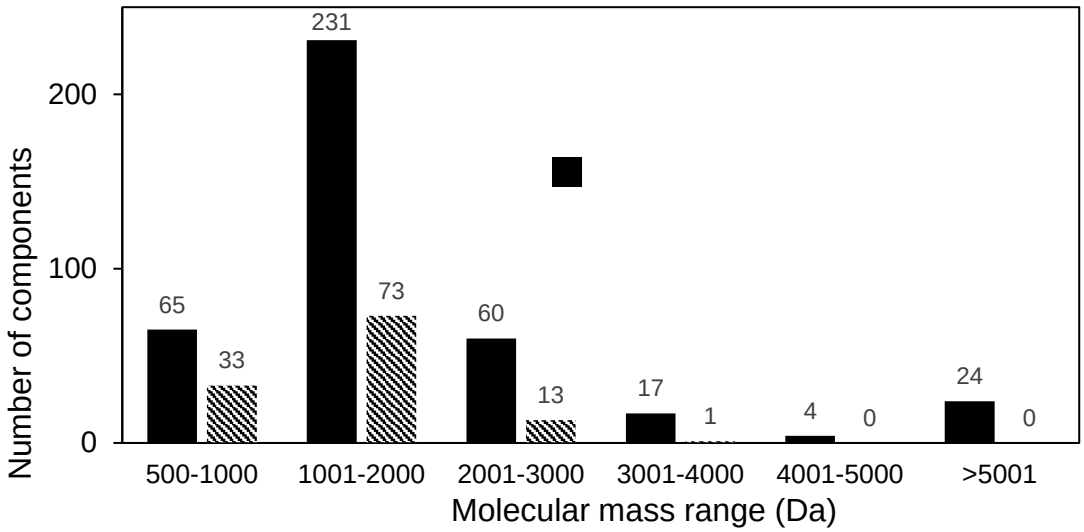


Figure 2-1. Molecular mass distribution of the components in the *L. poonaensis* venom.

Table 2-1. List of monoisotopic molecular masses (MM) of components in the *L. poonaensis* venom. Non-disulfide-bridged peptides are underlined.

MM	MM	MM	MM	MM	MM	MM	MM
512.30	936.54	1122.62	1319.76	1574.87	1806.10	2027.09	2815.37
601.35	955.59	1124.62	1321.70	1575.85	1809.01	2035.19	2901.47
614.38	956.50	1125.66	1329.79	1578.88	1815.00	2039.18	2934.74
636.36	968.63	1126.18	1330.83	1580.90	1816.03	2045.12	2953.62
652.33	970.53	1127.66	1331.73	1584.82	1820.07	2049.11	2959.51
670.45	974.62	1130.65	1335.75	1585.94	1822.96	2051.16	2983.16
675.04	976.46	1142.63	1336.73	1590.86	1839.0	2065.46	3029.28
688.40	978.59	1143.72	1343.73	1596.88	1840.07	2074.13	3078.69
690.40	980.46	1144.73	1344.82	1598.81	1841.40	2110.19	3093.82
715.40	982.53	1148.52	1349.73	1605.86	1844.97	2111.14	3094.71
727.39	987.50	1152.63	1358.76	1606.86	1845.83	2112.13	3109.71
729.48	988.53	1159.65	1390.74	1613.87	1848.88	2117.13	3193.70
732.38	989.52	1160.67	1392.75	1614.89	1852.90	2119.32	3210.95
741.40	991.60	1167.75	1396.78	1617.95	1853.91	2120.76	3295.93
747.42	997.66	1170.64	1401.86	1619.96	1857.02	2130.14	3318.92
752.42	1000.62	1171.62	1404.76	1623.88	1860.65	2133.92	3319.94
753.35	1001.61	1184.67	1408.78	1624.88	1866.00	2146.32	3326.83
757.47	1002.62	1187.66	1415.81	1633.96	1869.98	2147.31	3350.90
761.43	1012.57	1188.64	1416.81	1643.85	1873.04	2180.30	3392.07
786.40	1016.63	1196.70	1418.75	1644.94	1875.03	2181.92	3401.73
794.48	1019.55	1201.69	1426.77	1645.91	1877.12	2187.14	3407.73
798.53	1026.57	1205.71	1433.83	1660.95	1883.15	2194.27	3877.61
808.48	1031.65	1206.48	1434.83	1665.98	1898.97	2197.21	3878.55
809.45	1033.56	1211.61	1435.85	1667.93	1911.12	2216.16	4074.40
812.47	1034.26	1216.70	1444.80	1673.94	1927.08	2217.34	4143.37
816.49	1039.54	1226.83	1448.80	1693.99	1929.10	2223.28	4354.80
822.46	1046.63	1230.68	1452.82	1697.92	1930.14	2224.27	4526.20
826.49	1047.63	1231.76	1457.79	1698.86	1943.08	2251.33	5518.23
829.49	1049.61	1237.69	1461.83	1702.95	1948.16	2258.23	5538.23
832.52	1054.51	1238.80	1467.80	1706.98	1952.10	2259.23	5541.26
836.48	1056.68	1246.65	1472.89	1708.95	1954.14	2272.44	5579.55
840.47	1057.56	1247.75	1475.79	1711.92	1957.12	2281.25	5580.63
854.48	1058.54	1248.71	1483.90	1716.00	1958.06	2282.31	5582.52
856.47	1062.65	1251.70	1485.77	1716.89	1959.07	2337.12	5584.54
857.49	1063.61	1252.74	1501.90	1727.97	1969.11	2338.45	5585.58
860.45	1065.56	1255.70	1503.85	1728.96	1970.08	2341.27	5609.52
865.51	1071.58	1259.67	1506.88	1736.90	1972.09	2358.89	5614.53
871.51	1072.57	1266.20	1513.79	1741.08	1975.01	2370.34	5629.46
874.50	1073.65	1272.78	1514.96	1749.03	1977.10	2371.32	6128.72
875.51	1083.61	1273.71	1515.83	1764.89	1981.07	2500.41	6198.66
882.48	1087.58	1284.70	1519.82	1765.02	1982.13	2532.40	6205.61
884.52	1089.58	1286.77	1520.82	1773.99	1983.08	2533.34	6418.14
890.44	1095.60	1287.70	1532.85	1774.85	1987.13	2597.32	6732.03
895.46	1096.72	1297.74	1536.27	1780.95	1988.10	2606.16	6841.00
899.48	1097.54	1300.76	1545.86	1781.02	1989.67	2611.49	6978.47
904.45	1103.67	1301.74	1546.92	1782.97	1998.07	2615.24	7215.43
910.48	1110.75	1302.73	1556.85	1795.97	2011.09	2694.55	7357.29
911.61	1113.62	1313.71	1560.87	1796.98	2012.03	2715.45	7378.37
928.60	1116.64	1314.77	1562.89	1798.93	2017.15	2716.43	7379.37
932.53	1117.62	1315.76	1569.81	1799.01	2019.13	2811.48	7381.25

2.3.2 Isolation of non-disulfide-bridged peptides

The crude venom was separated into 9 fractions by RP-HPLC (Figure 2-2a). LC/MS analysis confirmed that no peptidic components were eluted after Fr. 9. Because of their relatively high hydrophobicity, AMPs in spider venom generally elute after most of the components in RP-HPLC analysis (Corzo et al., 2002; Santos D. M., et al., 2010; Wang et al., 2016). Therefore, we searched for an AMP candidate for non-disulfide-bridged peptides in Fr. 8 and Fr. 9. As observed in RP-HPLC analysis, there is no major peak in Fr. 8. Although the contribution of this fraction's components to the venom's activity may be limited, they are considered to be a source of bioactive peptides, such as AMPs. MALDI-TOF-MS analysis revealed that Fr. 8 contains several non-disulfide-bridged peptides (Figure 2-3). Two peptides with molecular masses of 2370 Da (lyp2370) and 1875 Da (lyp1875) were detected with high intensity. Fr.8 was further separated by RP-HPLC using a different gradient condition, and MALDI-TOF-MS analyzed each fraction. The result showed that peak A contains lyp2370 and lyp1875 (Figure 2-2b). In addition, we searched for other non-disulfide-bridged peptides detected at relatively low intensity. As a result, a peptide with a molecular mass of 2130 Da (lyp2130) was found in peak B, which gave an MS signal strong enough for MS/MS analysis. Thus, these two subfractions were subjected to MS/MS analysis. Next, Fr. 9 was analyzed by MALDI-TOF-MS, which revealed a peptide with a monoisotopic molecular mass of 1987 Da (lyp1987) as a major component (Figure 2-3). This fraction was subjected to MS/MS analysis without further purification.

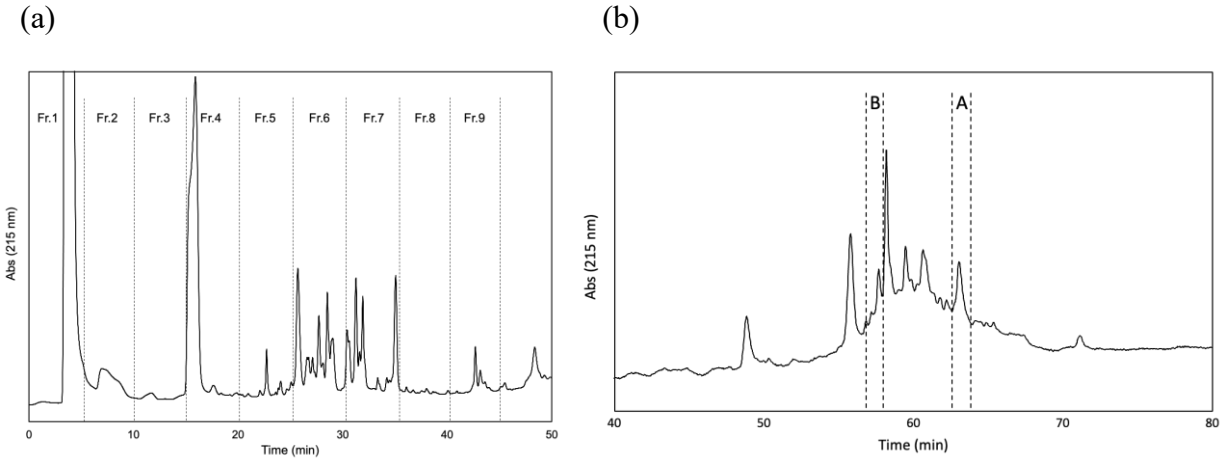


Figure 2-2. HPLC fractionation of the *L. poonaensis* venom. (a) HPLC separation of the crude venom using a C18 column. (b) Second HPLC purification of Fr.8 (35-40 min) using a different gradient condition.

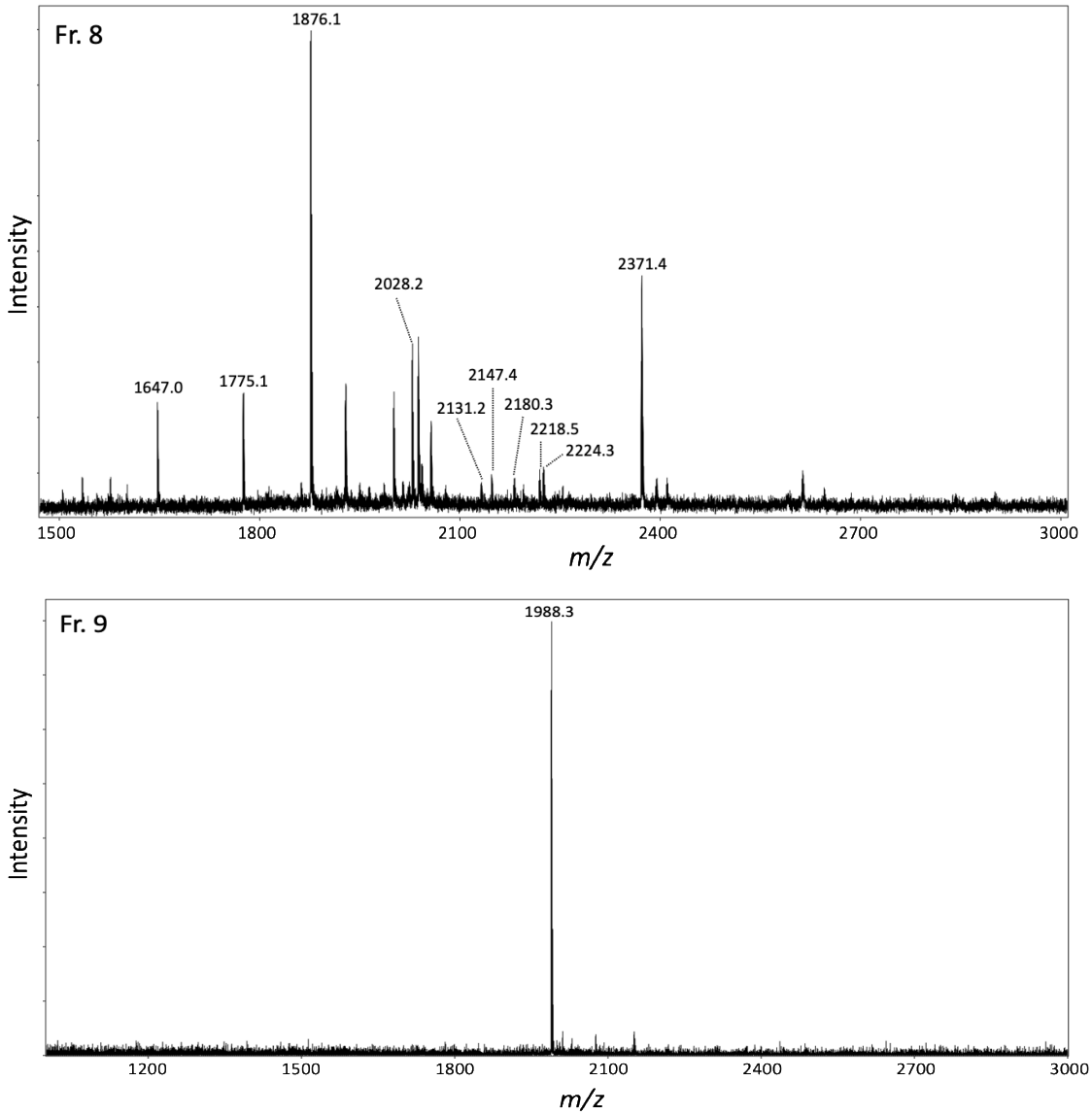


Figure 2-3. Mass spectra of Fr. 8 and Fr. 9 obtained by MALDI-TOF-MS. The *m/z* values were shown only for non-disulfide-bridged peptides listed in Table 2-1.

2.3.3 Structural determination of non-disulfide-bridged peptides

Product ion spectra of four non-disulfide-bridged peptides were obtained by MS/MS analysis using a MALDI-TOF/TOF mass spectrometer. When the production spectra were compared, similar fragment patterns were observed between lyp2370 and lyp2130 and lyp1987 and lyp1875 (Figure 2-4). This suggests that lyp2130 and lyp1875 consist of partial sequences of lyp2370 and lyp1987. Therefore, the sequences of lyp2370 and lyp1987 were further investigated. MS/MS analysis allowed us to estimate the sequence of lyp2370, as shown in (Figure 2-5). Leu and Ile residues were discriminated based on the side-chain fragmentation observed using a MALDI-TOF/TOF mass spectrometer, except for that at the C-terminus, for which no fragmentation is required Leu/Ile discrimination was observed (Figure 2-6).

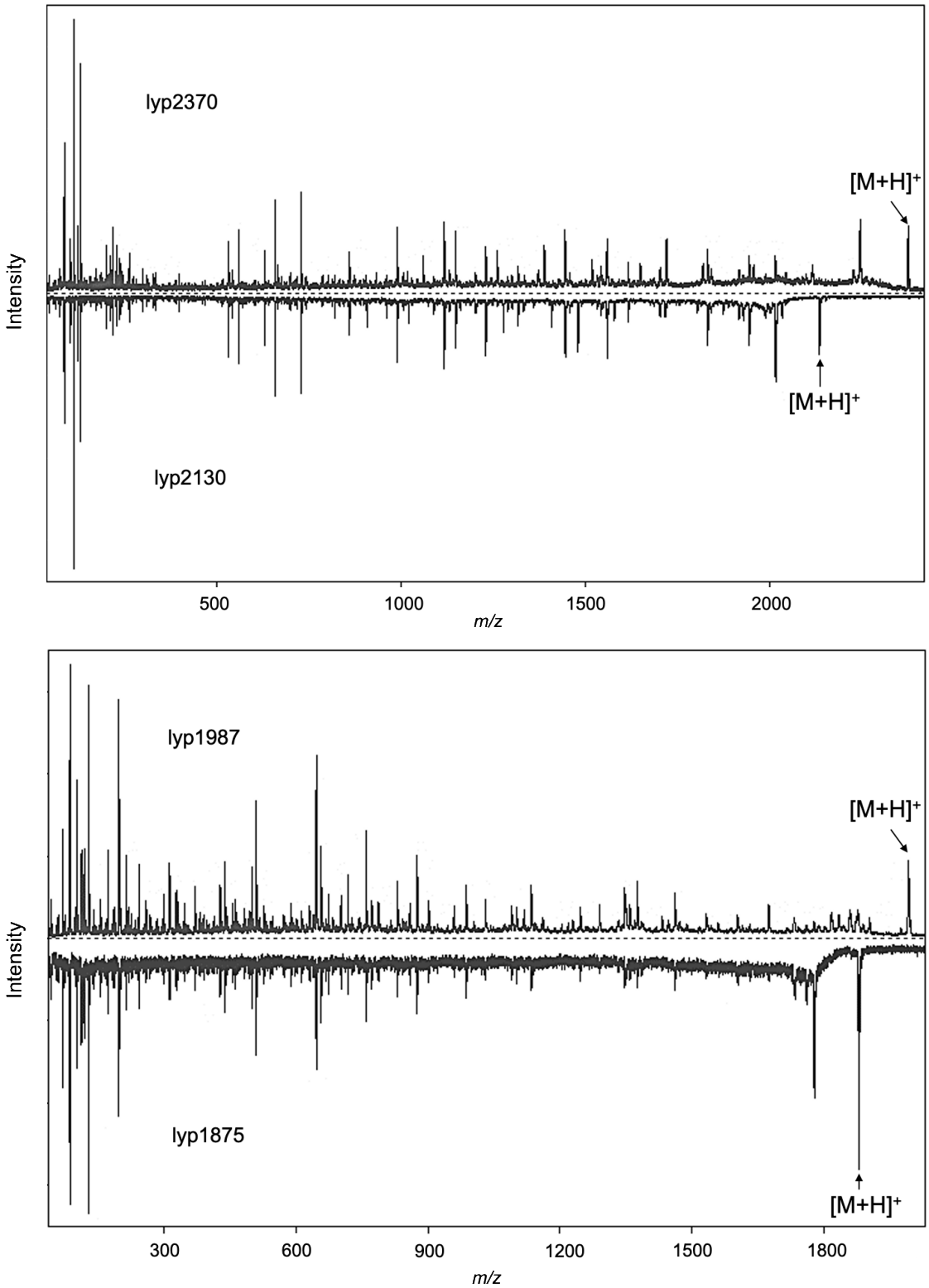


Figure 2-4. Comparisons of product ion spectra between lyp2370 and lyp2130, and between lyp1987 and lyp1875.

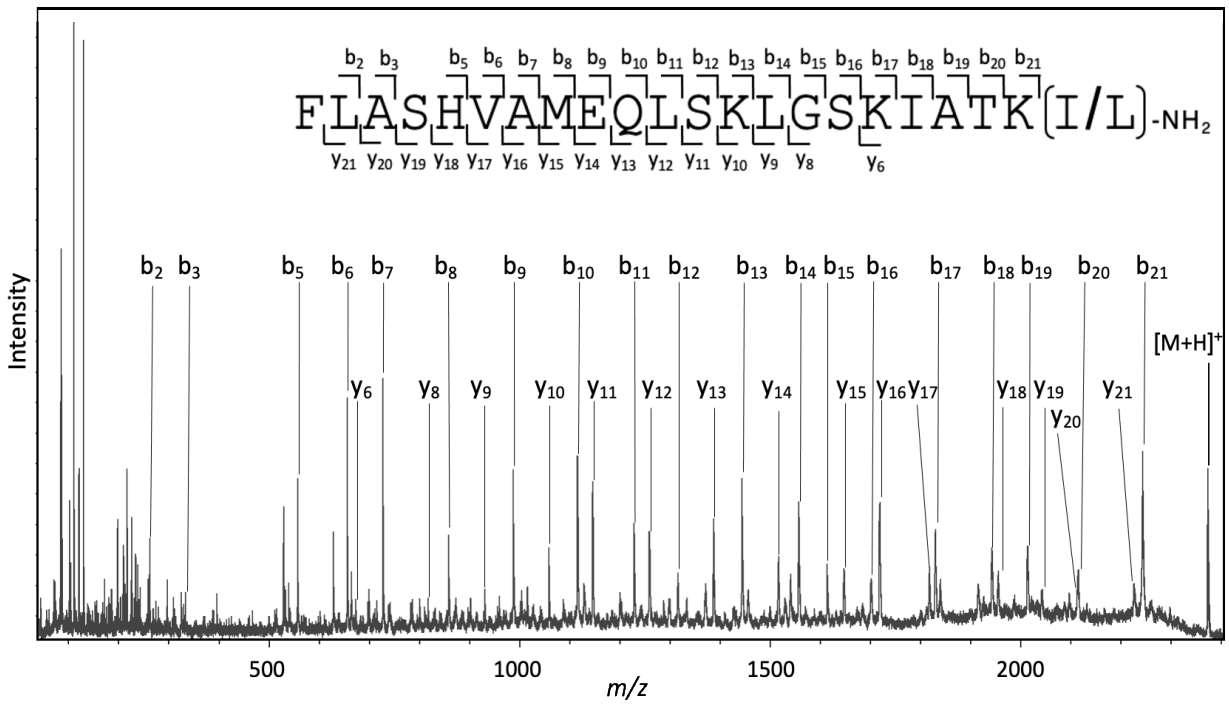


Figure 2-5. Product ion spectra of lyp2370 obtained by a MALDI-TOF/TOF mass spectrometer and the estimated sequences.

Discrimination between Lys and Gln residues was performed using an ESI-IT-TOF mass spectrometer, which provides data accurate enough for Lys/Gln discrimination (Table 2-2). Two candidate peptides were synthesized to determine whether the C-terminal residue is Leu or Ile, and their HPLC retention times were compared with that of the natural sample (Figure 2- 8). When the peptide having the C-terminal Leu residue was injected with the natural sample, they were detected as a single peak. On the other hand, the peptide having the C-terminal Ile residue was detected as a different peak from that of the natural sample. This result indicates that the C-terminal residue of lyp2370 is Leu. Since the molecular mass difference between lyp2370 and lyp2130 was 240.2, the latter peptide was expected to have a sequence lacking the two C-terminal residues (Lys-Leu-NH_2). The observed fragment ions of lyp2130 are in good agreement with those estimated from the sequence (Figure 2-4). A BLAST search showed that lyp2370 shares sequence similarity to M-lycotoxin-Ls4a, a peptide identified from the spider *Lycosa singoriensis* (Figure 2-7), although its biological activity has not been evaluated (Budnik, et al., 2004).

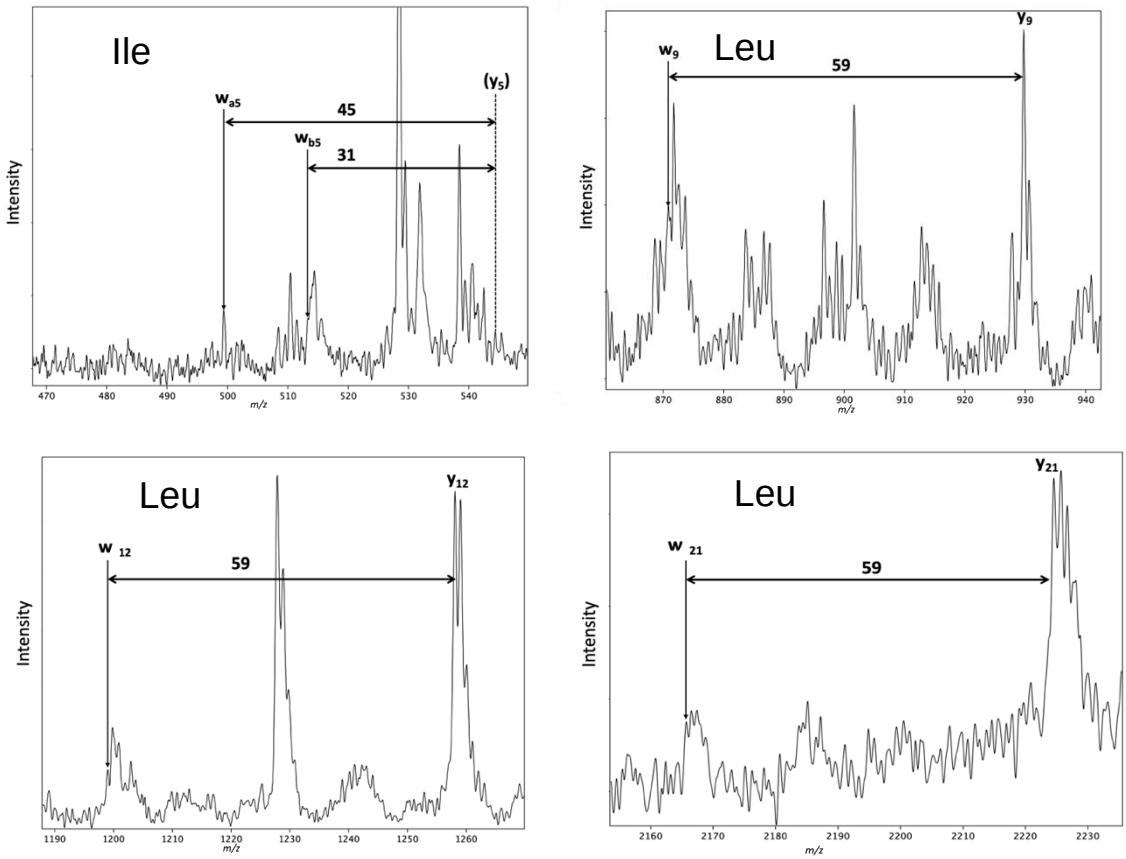


Figure 2-6. Discrimination between Leu/Ile in the sequence of lyp2370. Vertical arrows indicate observed w-ions.

Table 2-2. Comparison of the *m/z* values of fragment ions of lyp2370 between calculated and observed values.

FLASHVAME**Q**LSZLGS**K**LAT**K**X-NH_2

y_{20} y_{19} y_{18} y_{12} y_{11} y_4 y_3

Fragment	Charge	Calcd.	Obsvd.	Difference
y_3	1	360.261	360.245	0.016
y_4	1	431.298	431.287	0.011
y_{11}	2	572.874	572.867	0.007
y_{12}	2	629.416	629.403	0.014
y_{18}	3	651.715	651.703	0.012
y_{19}	3	680.726	680.713	0.013
y_{20}	3	704.405	704.398	0.007

lyp2370	FLASHVAMEQLSKLGSKIATKL-NH_2
lyp2130	FLASHVAMEQLSKLGSKIATK
M-lycotoxin-Ls4a	MXASHXAFEKXSKXGSKHTMX-NH_2
lyp1987	GRLQAFLAKMEKIAAQTL-NH_2
lyp1875	GRLQAFLAKMEKIAAQT
M-lycotoxin-Ls3b	GRLQAFLAKMEKIAAQTL-NH_2
M-lycotoxin-Ls3a	GKLQAFLAKMEKIAAQTL-NH_2

Figure 2-7. Sequences of non-disulfide-bridged peptides identified in this study and their multiple sequence alignment with similar peptides. X in the sequence indicates Leu or Ile residues.

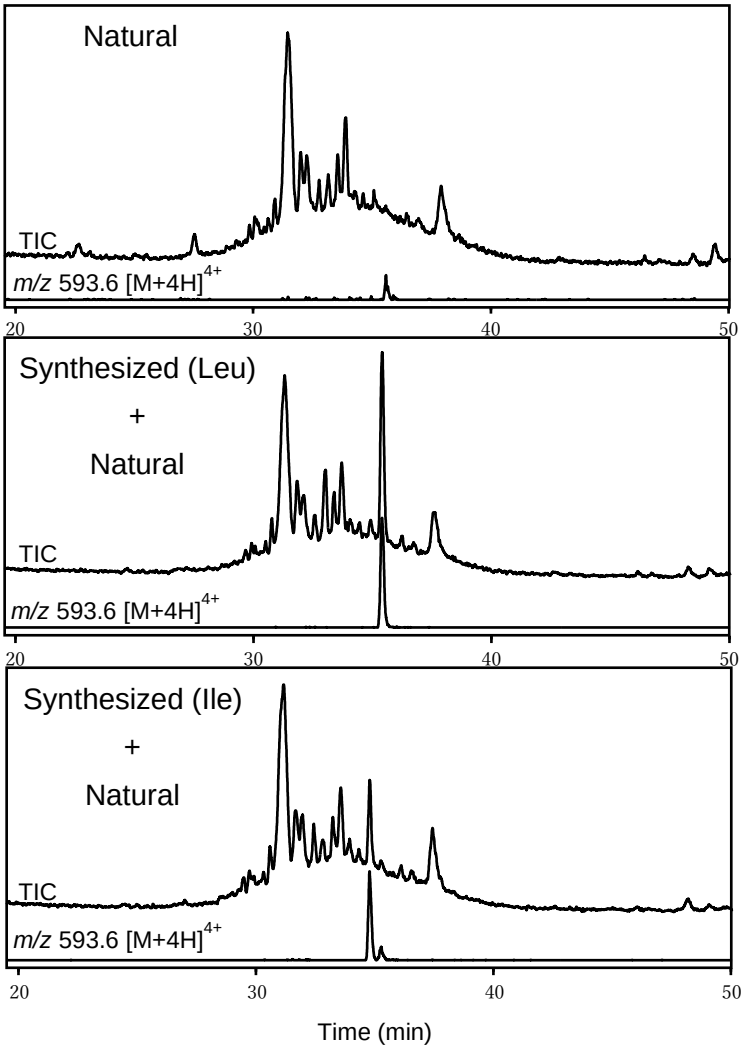


Figure 2-8. Comparison of HPLC retention times of lyp2370 between natural and synthesized samples. HPLC fractions containing lyp2370 were used as a natural sample. LC/MS analysis was performed for the natural sample alone and mixtures of the natural and synthesized samples.

Next, the MS/MS sequencing analysis of lyp1987 was performed using the same procedure used for lyp2370 (Figures 2-9 and 10). The sequence of lyp1987 was estimated as shown in Figure 2-9, except for the N-terminal two residues. Due to the absence of fragment ions, such as y_{17}, only the combination of two residues (Gly-Arg or Arg-Gly) could be determined for the N-terminal sequence (Figure 2-9, and Table 2-3). BLAST search using the deduced sequence revealed that lyp1987 has the same sequence as M-lycotoxin-Ls3b from the venom of *Lycosa singoriensis*, which has an N-terminal Gly-Arg sequence (Budnik, et al., 2004). The peptide was synthesized to confirm that lyp1987 has a structure identical to M-lycotoxin-Ls3b, and its HPLC retention time was compared with that of the natural sample (Figure 2-11).

When the peptide was injected with the natural sample, they were detected as a single peak, indicating that lyp1987 is identical to M-lycotoxin-Ls3b. It is known that the venom of closely related species occasionally contains peptides with the same sequence (Martin-Eauclaire and Bougis, 2012; Nguyen et al., 2014). The fact that the same or very similar peptides exist in different species may suggest that they significantly affect their venom. As observed for lyp2130, lyp1875 is likely to have a sequence lacking the C-terminal residue (Leu-NH_2) of lyp1987 based on the mass difference (112.0) between them (Figure 2-7). A detailed examination of the fragment ions in the production spectrum of lyp1875 confirmed this structure (Figure 2-3).

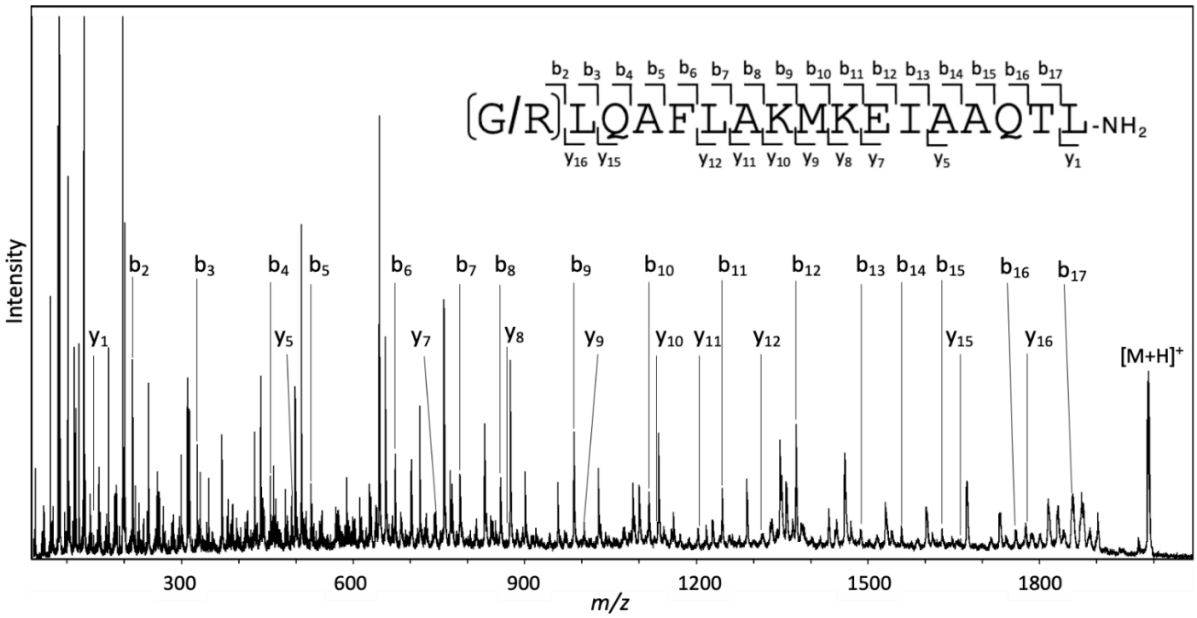


Figure 2-9. Product ion spectra of lyp1987 obtained by a MALDI-TOF/TOF mass spectrometer and the estimated sequences.

Table 2-3. Comparison of the *m/z* values of fragment ions of lyp1987 between calculated and observed values.

b_5 b_6 b_7 b_8 b_{10} b_{11} b_{12} b_{13} b_{14} b_{15} b_{16} b_{17}

(G/R)L**Q**AFLA**K**M**K**EIAA**Q**TL-NH_2

Fragment	Charge	Calcd.	Obsvd.	Difference
b_5	1	526.310	526.305	0.004
b_6	1	673.378	673.375	0.003
b_7	1	786.462	786.463	-0.001
b_8	1	857.499	857.497	0.003
b_{10}	2	558.821	558.813	0.008
b_{11}	2	622.868	622.855	0.013
b_{12}	2	687.390	687.378	0.011
b_{13}	2	743.932	743.929	0.003
b_{14}	2	779.450	779.450	0.001
b_{15}	2	814.969	814.968	0.001
b_{16}	2	878.998	878.996	0.002
b_{17}	2	929.522	929.525	-0.003

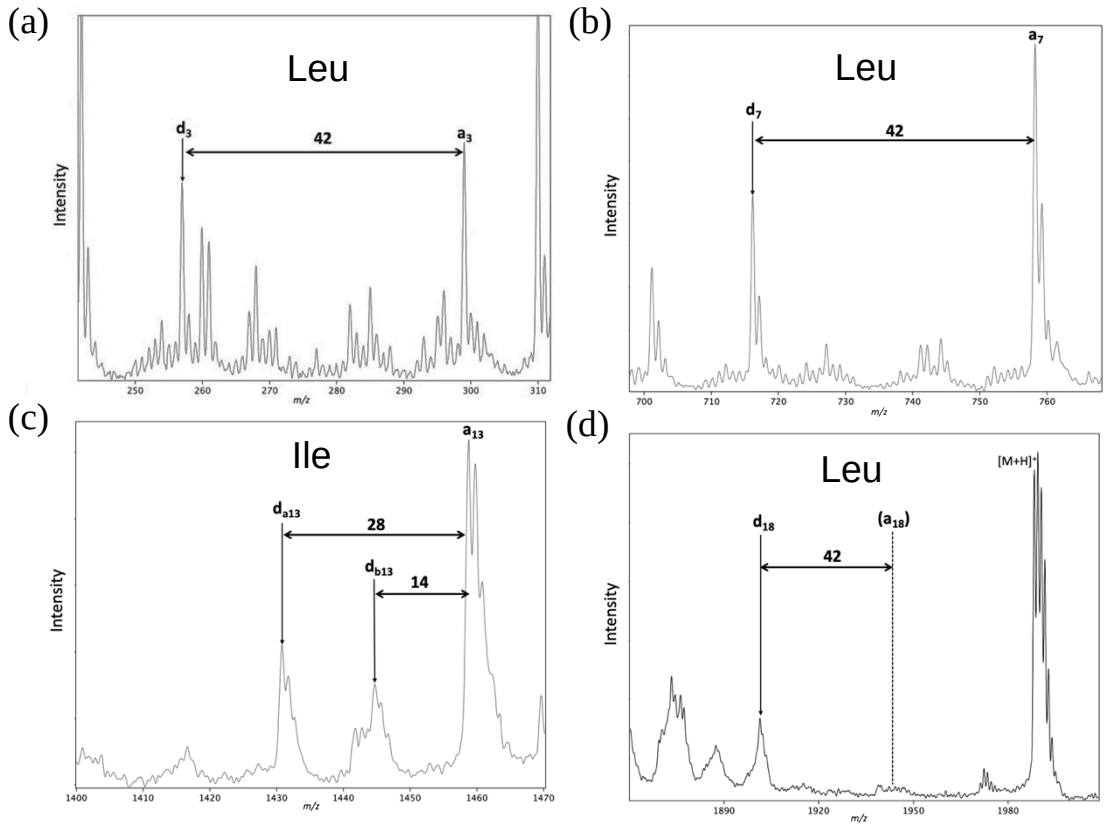


Figure 2-10. Discrimination between Leu/Ile in the sequence of lyp1987. Vertical arrows indicate observed d-ions.

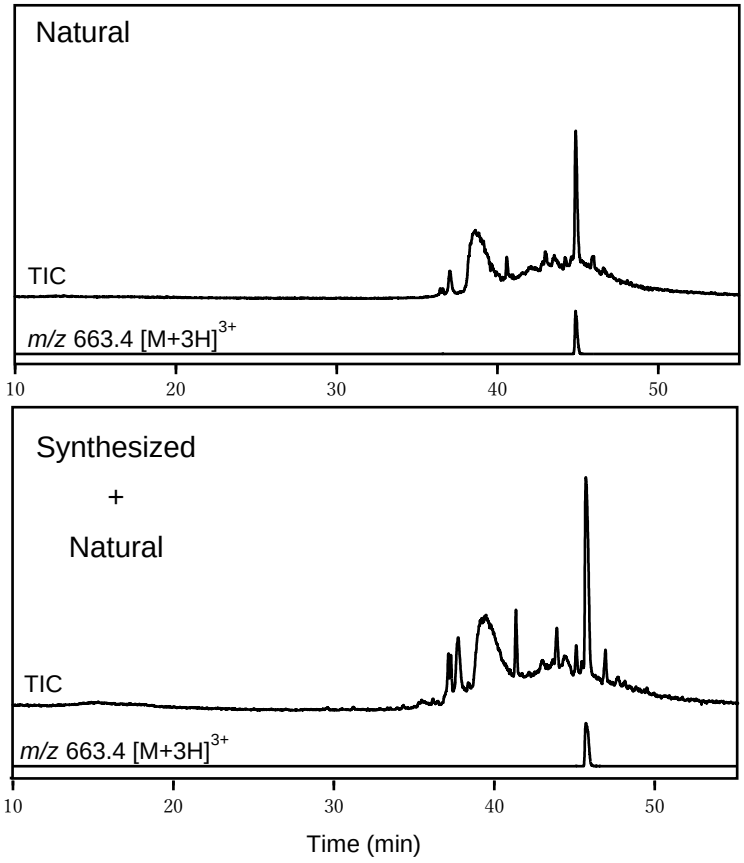


Figure 2-11. Comparison of HPLC retention times of lyp1987 between natural and synthesized samples. HPLC fractions containing lyp1987 were used as a natural sample. LC/MS analysis was performed for the natural sample alone and for mixtures of the natural and synthesized samples.

2.3.4 Biological activity

Antibacterial activity of synthesized lyp2370 and lyp1987 was evaluated against Gram-negative (*E. coli*) and Gram-positive (*S. aureus* and *B. subtilis*) bacteria. The activity of M-lycotoxin-Hc1a (IWLTALKFLGKHAAKHLAKQQLSKL-NH_2) identified from the Lycosidae spider *Hogna carolinensis* (Yans and Adams, 1998), was also evaluated for comparison. As shown in Table 2-4, lyp2370 showed only weak antibacterial activity against *B. subtilis*. On the other hand, lyp1987 showed activity higher than lyp2370 against three bacterial species, as previously observed for M-lycotoxin-Ls3b (Budnik, et al., 2004). M-lycotoxin-Hc1a showed higher activity than the other two peptides against three bacterial species. Since the antibacterial activity of lyp2370 is weak, its contribution to the venom activity may be marginal compared to that of lyp1987. However, we thought that comparing these peptides would provide information on structural factors essential for the expression of antibacterial activity.

It is known that an amphipathic α-helical structure is critical for the expression of antibacterial activity of AMPs (Avci, et al., 2018). To determine their secondary structure in solution, CD spectroscopy analysis was performed for lyp2370 and lyp1987 (Figures 2-12a and 12b). As shown in Table 2-5, both peptides adopt an α-helical structure in the presence of TFE, a secondary structure-promoting solvent. In addition, both peptides are likely to have an amphipathic nature based on the helical wheel projection (Figures 2-12c and 12d). Since the surface of bacteria is negatively charged, its electrostatic interaction with cationic sites of AMPs is essential in their antibacterial activity (Malanovic and Lohner, 2016). The relatively high activity of M-lycotoxin-Hc1a can be attributed to its high net charge (+6) (Jiang and Hodges, 2007). Since lyp2370 and lyp1987 have the same net charge (+3), the extremely low activity of lyp2370 is likely to be caused by factors other than the net charge. Both lyp2370 and lyp1987 contain three cationic residues (three Lys residues for lyp2370, two Lys residues, and one Arg residue for lyp1987) on the hydrophilic face of the α-helical structure with the different arrangement (Figures 2-12c and 12d). It is known that the position, arrangement, and combination of anionic and cationic amino acid residues on the hydrophilic face significantly affect the activity of AMPs (Dhople and Nagaraj, 2005). Taken together, the lyp2370 and lyp1987 low activity are due to a Glu residue at the hydrophilic face by interfering with its association with bacterial membranes.

To examine this hypothesis, the lower active peptide lyp2370 were subject to structural relation activity test, we synthesized lyp2370 analogs, in which the Glu residue was substituted with neutral Ala ([E9A]lyp2370) or basic Arg ([E9R]lyp2370), and measured their antibacterial activity (Table 2-4). The result showed that substitution with Ala resulted in an 8-fold increase in antibacterial activity against *B. subtilis*. Furthermore, a 16-fold increase in the activity against *B. subtilis* was observed after substitution with Arg. In this case, the peptide also became active against *E. coli*. These results clearly indicated that the low antibacterial activity of lype2370 is due to an anionic Glu residue at the central site of the hydrophilic face.

The analogs CD spectroscopy analysis was performed to determine the effect of neutral Ala and the basic Arg substitution on lyp2370 secondary structure in solution (Figures 2-13a and 13b). Both analogs still show α-helical structure in TFE with some difference in α-helix value Table 2-5. Also, both peptides still have an amphipathic nature based on the helical wheel projection (Figures 2-13c and 13d).

It is also possible that lyp2370 may have other functions in the venom from a biological point of view. Therefore, we evaluated insect toxicity and hemolytic activity of lyp2370 and analogs, lyp1987, and M-lycotoxin-Hc1a (Table 2-4). No insect toxicity was observed for lyp2370 or analogs, whereas lyp1987 and M-lycotoxin-Hc1a induced transient paralysis in crickets. As for hemolytic activity, none of the peptides including lyp2370 analogs showed activity. These results suggest that lyp2370 has unknown biological functions other than antibacterial, insecticidal, and hemolytic activities.

2.4 Summary

In summary, we analyzed the venom of the spider *L. poonaensis* using mass spectrometry to find non-disulfide-bridged peptides as an AMP candidate. The result revealed that 120 out of 401 components in the venom were non-disulfide-bridged peptides. The amino acid sequences of two peptides (lyp2370 and lyp1987) were determined by MS/MS analysis, confirmed by comparison with synthesized samples. We also found that two other peptides (lyp2130 and lyp1875) have C-terminally truncated lyp2370 and lyp1987, respectively. The antibacterial activity test revealed that lyp2370 has only weak activity, whereas lyp1987, which has a structure identical to M-lycotoxin-Ls3b, showed significant activity. The lower activity of lyp2370 was due to the presence of a Glu residue on the hydrophilic face of its amphiphilic α-helical structure.

Table 2-4. Biological activities of lyp2370, lyp2370 analogs, lyp1987, and M-lycotoxin-Hc1a.

Peptide	Antibacterial activity MIC (μM)			Insect toxicity ED_{50} (μg/g weight)	Hemolytic activity EC_{50} (μM)	Net Charge
	E. coli	*S. aureus*	*B. subtilis*			
lyp2370	>100	>100	100-200	>1000	>100	+3
[E9A]lyp2370	>100	>100	12.5-25	>1000	>100	+4
[E9R]lyp2370	25-50	>100	6.25-12.5	>1000	>100	+5
lyp1987 (M-lycotoxin-Ls3b)	15-20	75-100	2.5-5	146	>100	+3
M-lycotoxin-Hc1a[a)]	2.5-5	1-2.5	0.5-1	58	>100	+6

a) Sequence; IWLTALKFLGKHAAKHLAKQQLSKL-NH_2

Table 2-5. Estimated secondary structures included in lyp2370, lyp2370 analogs, and lyp1987.

Peptides	Solvents [a)]	Percentage of secondary structure [b)]			
		α-helix	β-sheet	Turns	Unordered
lyp2370	I	8	31	24	35
	II	74	7	5	14
[E9A]lyp2370	I	11	30	26	33
	II	47	13	14	27
[E9R]lyp2370	I	10	30	27	33
	II	29	21	20	29
lyp1987	I	10	36	21	32
	II	63	9	7	20

[a)] Solvent I indicate 0.2 M phosphate buffer (pH7.0), and solvent II indicates 50% TFE in 0.2 M phosphate buffer (pH7.0). [b)] Calculated using the DichroWeb server with the CDSSTR method and reference data set 4.

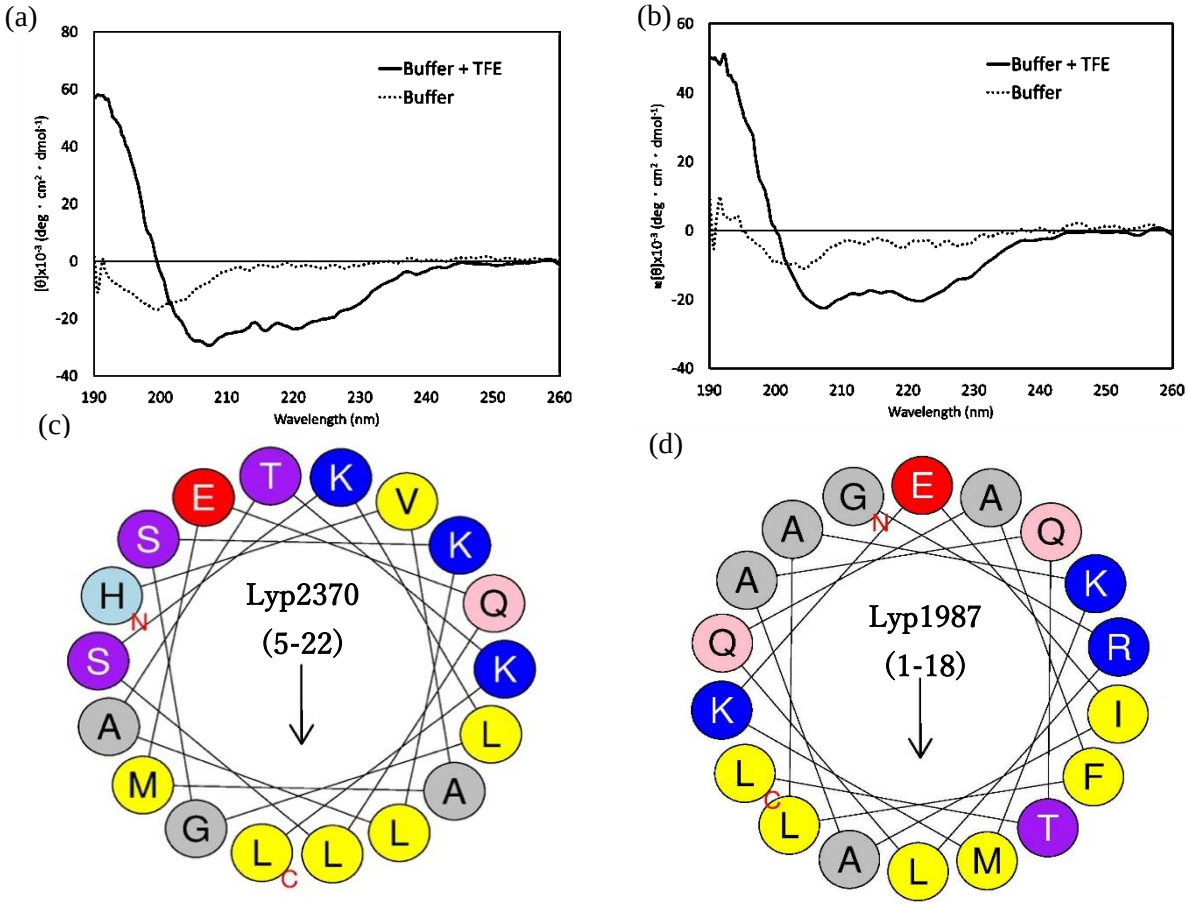


Figure 2-12. Secondary structure analysis. CD spectra of lyp2370 (a) and lyp1987 (b) in 0.2 M phosphate buffer at pH 7.0 (dotted line) and in 50% TFE in 0.2 M phosphate buffer at pH 7.0 (solid line). Helical wheel projections of lyp2370 (c) and lyp1987 (d). Arrows indicate hydrophobic moment.

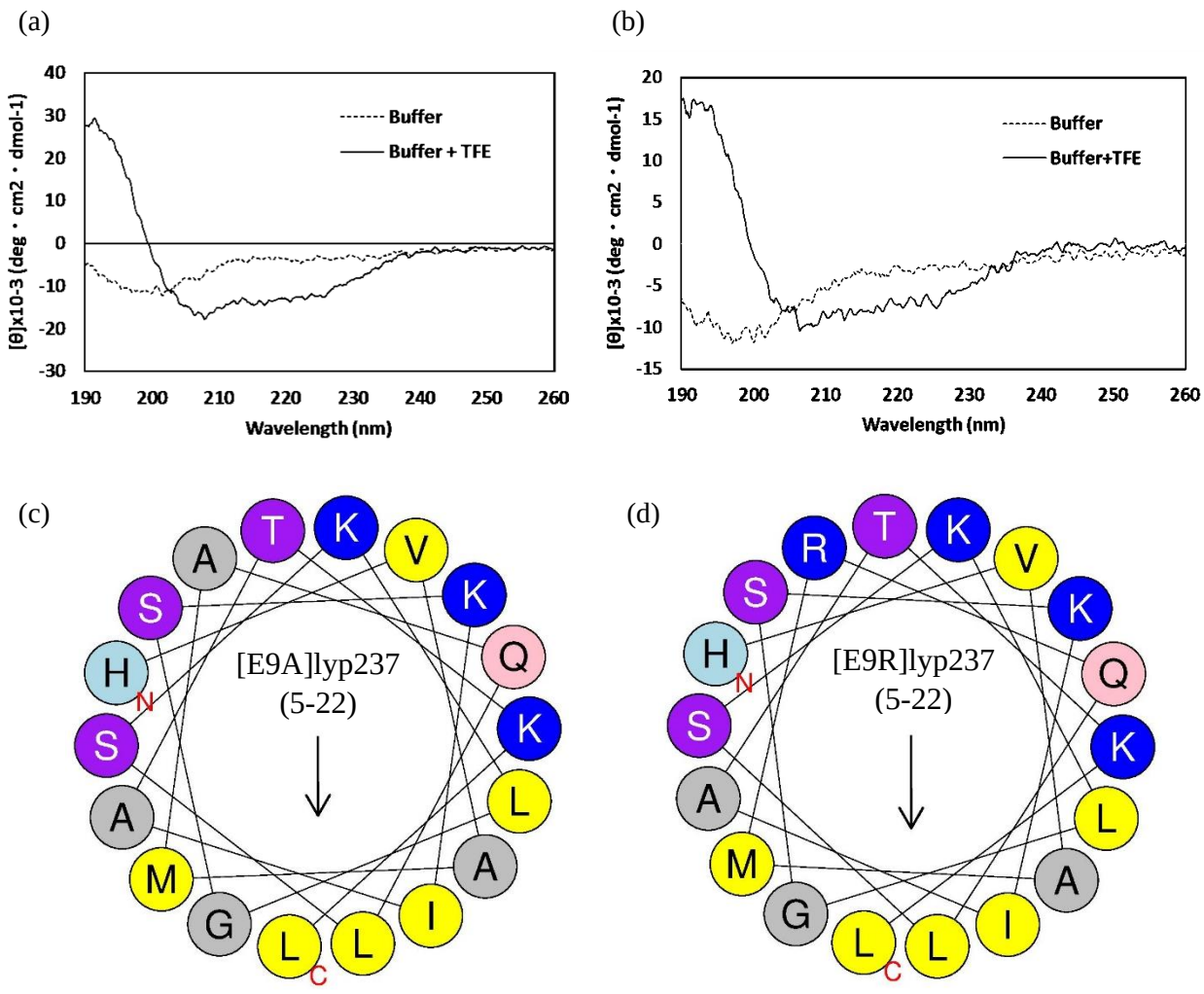


Figure 2-13. Secondary structure analysis. CD spectra of lyp2370 analogs (a) [E9A]lyp237 and [E9R]lyp237 (b) in 0.2 M phosphate buffer at pH 7.0 (dotted line) and in 50% TFE in 0.2 M phosphate buffer at pH 7.0 (solid line). Helical wheel projections of [E9A]lyp237 (c) and [E9R]lyp237(d). Arrows indicate hydrophobic moment.

References

Akef H.M. Anticancer, antimicrobial, and analgesic activities of spider venoms. *Toxicol Res* 2018;7;381-95.

Ashfaq M., Blagoev G., Tahir H.M., Khan A.M., Mukhtar M.K., Akhtar S., Butt A., Mansoor S., Hebert P.D.N. Assembling a DNA barcode reference library for the spiders (Arachnida: Araneae) of Pakistan. *Plos One* 2019;14;ARTN e0217086.

Avci F.G., Akbulut B.S., Ozkirimli E. Membrane active peptides and their biophysical characterization. *Biomolecules* 2018;8;ARTN 77.

Budnik B.A., Olsen J.V., Egorov T.A., Anisimova V.E., Galkina T.G., Musolyamov A.K., Grishin E.V., Zubarev R.A. De novo sequencing of antimicrobial peptides isolated from the venom glands of the wolf spider *Lycosa singoriensis*. *J Mass Spectrom* 2004;39;193-201.

Corzo G., Villegas E., Gomez-Lagunas F., Possani L.D., Belokoneva O.S., Nakajima T. Oxyopinins, large amphipathic peptides isolated from the venom of the wolf spider *Oxyopes kitabensis* with cytolytic properties and positive insecticidal cooperativity with spider neurotoxins. *J Biol Chem* 2002;277;23627-37.

Dhople V.M., Nagaraj R. Conformation and activity of delta-lysin and its analogs. *Peptides* 2005;26;217-25.

Jiang Z., Hodges R.S. Effects of net charge and the number of positively charged residues on the biological activity of amphipathic-helical cationic antimicrobial peptides. *Biopolymers* 2007;88;615-.

Kaiser E., Colescott R.L., Bossinger C.D., Cook P.I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* 1970;34;595-8.

Kuhn-Nentwig L., Stocklin R., Nentwig W. Venom composition and strategies in spiders: Is everything possible? *Adv Insect Physiol* 2011;40;1-86.

Langenegger N., Nentwig W., Kuhn-Nentwig L. Spider venom: Components, modes of action, and novel strategies in transcriptomic and proteomic analyses. *Toxins* 2019;11;ARTN 611.

Lozano-Fernandez J., Carton R., Tanner A.R., Puttick M.N., Blaxter M., Vinther J., Olesen J., Giribet G., Edgecombe G.D., Pisani D. A molecular palaeobiological exploration of arthropod terrestrialization. *Philos T R Soc B* 2016;371;ARTN 20150133.

Malanovic N., Lohner K. Antimicrobial peptides targeting gram-positive bacteria. *Pharmaceuticals* 2016;9;ARTN 59.

Martin-Eauclaire M.F., Bougis P.E. Potassium channels blockers from the venom of *Androctonus mauretanicus mauretanicus*. *J. Toxicol.* 2012;2012.

Nguyen B., Le Caer J.P., Araoz R., Thai R., Lamthanh H., Benoit E., Molgo J. Isolation, purification and functional characterization of alpha-BnIA from *Conus bandanus* venom. *Toxicon* 2014;91;155-63.

Santos D.M., Reis P.V., Pimenta A.M.C. Antimicrobial peptides in spider venoms, In: Gopalakrishnakone P., Corzo G.A., de Lima M.E., Diego-García E. (Eds). Spider Venoms. Dordrecht: Springer Netherlands, 2016. 361-77.

Santos D.M., Verly R.M., Pilo-Veloso D., de Maria M., de Carvalho M.A.R., Cisalpino P.S., Soares B.M., Diniz C.G., Farias L.M., Moreira D.F.F., Frezard F., Bemquerer M.P., Pimenta A.M.C., de Lima M.E. LyeTx I, a potent antimicrobial peptide from the venom of the spider *Lycosa erythrognatha*. *Amino Acids* 2010;39;135-44.

Strohalm M., Kavan D., Novak P., Volny M., Havlicek V. mMass 3: A cross-platform software environment for precise analysis of mass spectrometric data. *Anal Chem* 2010;82;4648-51.

Tang X., Yang J., Duan Z., Jiang L., Liu Z., Liang S. Molecular diversification of antimicrobial peptides from the wolf spider *Lycosa sinensis* venom based on peptidomic, transcriptomic, and bioinformatic analyses. *Acta Biochim Biophys Sin* 2020;52;1274-80.

Wang Y.J., Wang L., Yang H.L., Xiao H.L., Farooq A., Liu Z.H., Hu M., Shi X.L. The Spider Venom Peptide Lycosin-II Has Potent Antimicrobial Activity against Clinically Isolated Bacteria. *Toxins* 2016;8.

Whitmore L., Wallace B.A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* 2008;89;392-400.

Windley M.J., Herzig V., Dziemborowicz S.A., Hardy M.C., King G.F., Nicholson G.M. Spider-venom peptides as bioinsecticides. *Toxins* 2012;4;191-227.

Wu T., Wang M., Wu W.F., Luo Q.X., Jiang L.P., Tao H., Deng M.C. Spider venom peptides as potential drug candidates due to their anticancer and antinociceptive activities. *J Venom Anim Toxins* 2019;25;ARTN e146318.

Wullschleger B., Nentwig W., Kuhn-Nentwig L. Spider venom: enhancement of venom efficacy mediated by different synergistic strategies in *Cupiennius salei*. *J Exp Biol* 2005;208;2115-21.

Yans L.Z., Adams M.E. Lycotoxins, antimicrobial peptides from venom of the wolf spider *Lycosa carolinensis*. *J Biol Chem* 1998;273;2059-66.

CHAPTER 3

De novo transcriptome analysis of the venom gland of *L. poonaensis*

3.1 Introduction

Studies of spider venom started in the 20th century. The first studies were conducted to evaluate the effect of whole venom on animals and humans, which are useful for development of antivenin. After that, with the advent of separation technologies, such as high-performance liquid chromatography (HPLC), identification of the active components of the venom responsible for the toxicity was conducted. In these experiments, protein sequencing using Edman degradation was initially used to identify the amino acid sequences(Kuhn-Nentwig et al., 2002). Although this sequencing method is highly reliable, it requires a relatively large amount of samples and is time-consuming. This problem was largely overcome with the development of the sequencing technique using a tandem mass spectrometer, which dramatically accelerated the venom research due to its high sensitivity and throughput(Kuhn-Nentwig, et al., 2002).

However, even with analysis by these techniques, the speed of identification of active components is still limited. The recent advent of high-throughput DNA sequencing techniques has made it possible to obtain comprehensive amino acid sequences of peptides and proteins expressed in a given tissue (transcriptome), such as a venom gland, very quickly. Transcriptome analysis allows for fast and sensitive identification of venom components with limited sample availability(Kuhn-Nentwig, 2021; Kuhn-Nentwig, Schaller, & Nentwig, 2004). To date, transcriptome analysis of the venom gland has been performed in several spider species(Kuhn-Nentwig, 2021). However, the sequences obtained by this analysis contains signal and propeptide regions, which are often difficult to predict. Therefore, the combination of the information obtained by transcriptome analysis with those by mass spectrometric analysis is necessary to determine the mature region, which can dramatically improve the reliability of the structural information of venom components. In this study, de novo transcriptome analysis of the venom gland of *L. poonaensis* was performed to estimate the structure of all peptides and proteins expressed in the venom.

3.2 Material and methods

3.2.1 Biological materials

The *L. poonaensis* spiders were collected from the Western Desert of Egypt, as mentioned in the previous chapter, and delivered to the lab individually in paper boxes. The spiders were anesthetized on ice, and dissected at room temperature in distilled water. Venom glands were excised and directly placed in a microtube containing RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) at room temperature for 1 h, followed by storage at -80°C until RNA extraction. During transportation from Egypt to Japan, the samples were surrounded by ice to prevent a rise in temperature.

3.2.2 RNA extraction

Venom glands were homogenized manually in a glass tube of a microtissue grinder. The total RNA was extracted using the TRIzol reagent (Thermo Fisher Scientific) and further purified using the RNeasy Mini Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's instructions. The concentration and purity of total RNA samples were estimated using the Nanodrop Lite Spectrometer (Thermo Fisher Scientific).

3.2.3 Library preparation and sequencing

Library preparation and sequencing analysis were conducted at Bioengineering Lab. Co., Ltd. (Sagamihara, Japan). The concentration of the total RNA was measured using Quantus Fluorometer with QuantiFluor RNA system (Promega Madison, WI, USA). The quality of RNA was then analyzed using the 5200 Fragment Analyzer System (Agilent Technologies), Santa Clara, CA, USA). RNA sequencing libraries were prepared using the MGIEasy RNA Directional Library Prep Set (MGI Tech, Shenzhen, China) according to the manufacturer’s instructions. The concentration of the prepared library solution was determined using Qubit 3.0 Fluorometer with the dsDNA HS Assay Kit (Thermo Fisher Scientific). The quality of the library was analyzed using the 5200 Fragment Analyzer System with the dsDNA 915 Reagent Kit (Agilent Technologies). Single-stranded circular DNA was prepared using the constructed library and the MGIEasy Circularization Kit (MGI Tech) and DNA nanoballs (DNBs) were prepared using the DNBSEQ-G400RS High-throughput Sequencing Kit (MGI Tech). The 200 bp paired-end sequencing was performed on the DNBSEQ-G400 (MGI Tech).

3.2.4 De novo assembly and functional annotations

After adapter-trimming using cutadapt (1.9.1), the reads were further trimmed using sickle (1.33) with a minimum window quality score of 20. The reads shorter than 40 bp after trimming were removed. The high-quality reads were assembled into contigs using Trinity (2.10.0). Coding regions were predicted by TransDecoder (5.5.0). The redundant sequences were removed by clustering using CD-HIT at the threshold of 1.00. The predicted peptide sequences were submitted to homology searches against a local blast database constructed using sequences identified from spider venom (the UniProt Animal Toxin Annotation Project) to annotate the functions of identified peptides and proteins. For each annotated sequence, a BLAST search was individually performed against database with the class Arachnida (taxid: 6854). All annotated sequences were manually inspected to exclude internal and partial sequences. The complete sequences were submitted to SignalP (5.0) to determine the signal peptide sequence. Mature regions of linear peptides were predicted based on the presence of processing quadruplet motif (PQM) and inversed processing quadruplet motif (iPQM), which were identified by (Kozlov et al., 2006). Multiple sequence alignments were performed manually or by Clustal Omega (Madeira et al., 2019).

3.2.5 Nomenclature of peptides

The nomenclature of linear cytolytic peptides (M-Lycotoxin-Lpn) was largely the method proposed by King *e4t al*., where the M refers to the action target of the peptide (membrane), followed by a generic name (lycotoxin) based on taxonomic family. Lp refers to the first two characters of species name (*Lycosa poonaensis*) followed by a number (n) to distinguish similar toxins within them.

3.3 Results and Discussion

3.3.1 Transcriptome analysis of the venom gland of L. poonaensis.

A total of 33,055,014 raw reads were obtained by sequencing. After adapter- and quality-trimming, the clean reads (32,712,326) were assembled in a de novo fashion using Trinity, which resulted in 280,691 contigs with N50 of 743 bp (186–34,812 bp). BLASTP search was performed for predicted coding regions against a database consisting of reported sequences identified from spider venom. This resulted in the identification of 87 transcripts encoding peptides and proteins similar to those of other spider venoms as shown in Table S3-1. The number of sequences identified as neurotoxins was highest, followed by peptidases, cytolytic toxins, and other enzyme proteins (Figure 3-1).

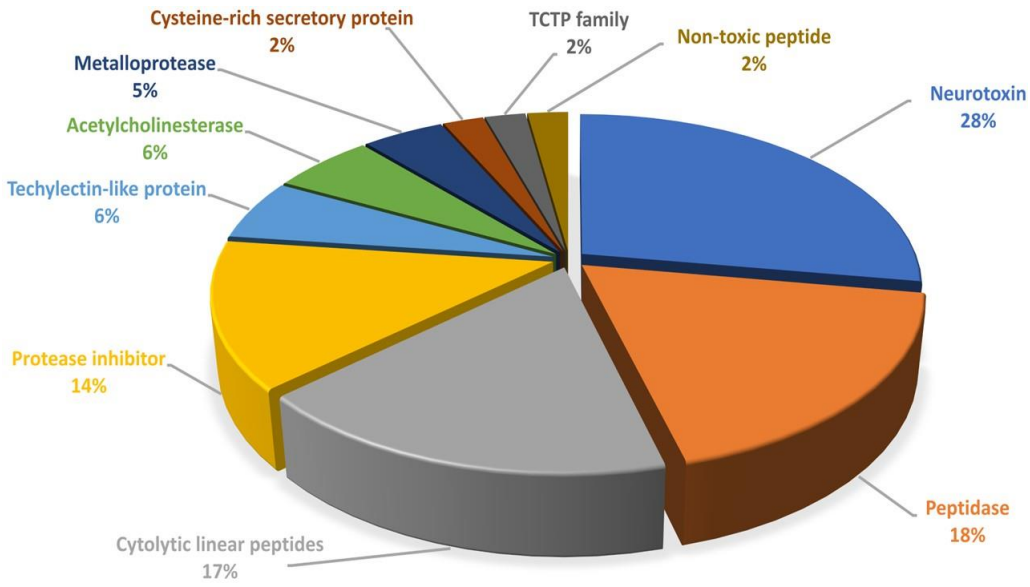


Figure 3-1. Proportion of the components identified by transcriptome analysis of the venom gland of *L. poonaensis*.

3.3.1.1 Neurotoxin

Neurotoxins are main active components of the spider venom. In this study, 24 contigs have sequences similar to the peptides that are expected to have neurotoxic effects. Among them, 3 contigs shared sequence similarity to the spider venom peptides that were experimentally confirmed to act on ion channels. The contig TRINITY_m.59186 has the sequence similar to Omega-ctenitoxin-Pn3a, which was identified from *Phoneutria nigriventer*. Omega-ctenitoxin-Pn3a irreversibly blocked calcium currents generated by Cav2.1 and Cav2.2 channels, whereas it showed partial and reversible inhibition on that by Cav2.3 channel(Cassola et al., 1998). The contig TRINITY_m.1192 shows similarity to Purotoxin-1 identified from *Alopecosa marikovskyi*. Purotoxin-1 has potent and selective inhibitory action on P2X3 receptors that are expressed in mammalian sensory neurons(Grishin et al., 2010; Kabanova et al., 2012). Other contigs also showed similarity to spider venom peptides. However, since functions of these peptides are still unknown, it is difficult to predict the activity of the peptides identified in this study based on the similarity.

3.3.1.2 Cytolytic toxin

Cytolytic toxins, which have a linear structure without disulfide bonds, play an important role in the venom of some spider species by showing various biological activities, such as insecticidal, antimicrobial, and hemolytic activities(Kuhn-Nentwig, et al., 2002).

In addition, some linear peptides were shown to enhance the insecticidal activity of other neurotoxins(Kuhn-Nentwig, et al., 2002; Kuhn-Nentwig, et al., 2004). In this study, 15 contigs contain sequences similar to cytolytic linear peptides.

The existence of the sequences identified in chapter 2 was confirmed in the several contigs. The sequence of lyp2370 was found in one contig (TRINITY_m.2773), whereas that of lyp1987 was found in 6 contigs (TRINITY_m.2667, TRINITY_m.2650, TRINITY_m.2660, TRINITY_m.2706, TRINITY_m.2678, and TRINITY_m.2773). As observed for linear peptides in other spider venom(Kozlov, et al., 2006), the sequence of lyp2370 and lyp1987 were repeatedly present in one single contig in some cases. Sequences of both lyp2370 and lyp1987 are also observed in the same contig. This type of precursor structure, which contains multiple mature regions, will be discussed in more detail in the following section. In addition to lyp2370 and lyp1987, 8 contigs (TRINITY_m.2649, TRINITY_m.2799, TRINITY_m.2698, TRINITY_m.2781, TRINITY_m.2667, TRINITY_m.2650, TRINITY_m.2660, and TRINITY_m.2706) and 4 contigs (TRINITY_m.2678, TRINITY_m.2734, TRINITY_m.2795, and TRINITY_m.2773) also contain sequences similar to LyeTx 1 and M-lycotoxin-Hc2a, respectively. LyeTx 1, which was identified from *Lycosa erythrognatha*, shows antimicrobial activity against bacteria and yeast as well as hemolytic activity(Santos et al., 2010). M-lycotoxin-Hc2a, which identified from *Hogna carolinensis*, shows various biological activity, such as antimicrobial, cytolysis, hemolysis, as well as neurotoxic action by disrupting membrane structure of prey organisms(Adao, Seixas, Gomes, Pessoa, & Bastos, 2008; Yans & Adams, 1998). These results indicate that there are other cytolytic linear peptides besides lyp2370 and lyp1987 in the *L. poonaensis* venom.

3.3.1.3 Other peptides

Protease inhibitor peptides are often observed in the spider venom. Although its function remains unknown, it is proposed that these peptides could protect the integrity of other venom components through their inhibitory activity on protease(Wan et al., 2013). In this study, 6 contigs have sequences similar to protease inhibitors. Four contigs (TRINITY_m.201956, TRINITY_m.221752, TRINITY_m.248055 and TRINITY_m.250351) have sequences similar to Kunitz-type protease inhibitor peptides. Two contigs (TRINITY_m.204789 and TRINITY_m.217252) show similarity to cysteine proteinase inhibitor and Kazal-type serine protease inhibitor, respectively(Undheim et al., 2013). Among them TRINITY_m.250351 shows similarity to kappaPI-theraphotoxin-Hs1a identified form *Cyriopagopus schmidti*. Interestingly, this toxin has dual function as a trypsin inhibitor and a potassium channel blocker(Sade et al., 2012; Yuan et al., 2008).

Translationally controlled tumor protein homologs (TCTPs) are known as histamine releasing factors, which promotes allergic response in mammalian tissues by inducing the release of histamine from basophiles or mast cells(Sade, et al., 2012).

Although TCTPs are observed in venom of some spider species, its biological role as a venom component in prey capture is still unknown. In this study, two contigs (TRINITY_m.7640 and TRINITY_m.24068) have sequences similar to TCTPs.

In addition, 8 contigs (TRINITY_m.41663, TRINITY_m.1940, TRINITY_m.41662, TRINITY_m.11922, TRINITY_m.240217, TRINITY_m.11914, TRINITY_m.106956, TRINITY_m.7591) show similarity to toxin-like peptides, whose functions are not yet understood.

3.3.1.4 Enzyme and protein

Spider venoms are known to contain enzymes and proteins, which can be classified into three functional classes; (1) enzymes acting as a spreading factor, (2) proteins with a function in the venom gland including the maturing of toxins, and (3) proteins directly acting on targets of the prey organism(Kuhn-Nentwig, Stocklin, & Nentwig, 2011). The biological roles of venom proteins are in most cases not experimentally confirmed, its function has been predicted based on its similarity to other known peptides.

The enzymes acting as a spreading factor can degrade or modify the extracellular matrix or the membrane of cells, facilitating the movement of toxins in the prey. The main group of enzymes responsible for this function are matrix-metalloproteases(Kuhn-Nentwig, et al., 2011; Langenegger, Nentwig, & Kuhn-Nentwig, 2019). In this study, 4 contigs (TRINITY_m.136344, TRINITY_m.44881, TRINITY_m.35058, TRINITY_m.220610) showed sequence similarity to Astacin-like metalloprotease identified from *Loxosceles intermedia*.

Peptidases play a critical role in the venom to produce mature peptides from precursors by recognizing the specific motif(Langenegger et al., 2018). In this study, 16 contigs (TRINITY_m.1011, TRINITY_m.15154, TRINITY_m.19936, TRINITY_m.9580, TRINITY_m.19952, TRINITY_m.60402, TRINITY_m.28887, TRINITY_m.43613, TRINITY_m.252247, TRINITY_m.77294, TRINITY_m.183905, TRINITY_m.183910, TRINITY_m.4971, TRINITY_m.4960, TRINITY_m.61171, TRINITY_m.71477) were identified as a peptidase. It is possible that some of these peptidases are involved in the maturation of linear peptides, which will be discussed later.

Acetylcholinesterases are also known to exist in some spider venoms, which possibly act on synaptic targets of the prey organism(Undheim, et al., 2013), and 5 contigs (TRINITY_m.101805, TRINITY_m.126349, TRINITY_m.229987, TRINITY_m.130212, and TRINITY_m.130210) showed similarity to this enzyme. Tachylectin-like proteins are hypothesized to be involved in protection of the venom gland against microbial infections(Kuhn-Nentwig, Langenegger, Heller, Koua, & Nentwig, 2019), and 5 contigs

(TRINITY_m.72849, TRINITY_m.257368, TRINITY_m.72846, TRINITY_m.72851, TRINITY_m.265263) showed similarity to this protein.

In addition to these proteins, cysteine-rich secretory proteins (CRISPs), which are widely distributed among various animal venoms, although their functions remain unclear(Langenegger, et al., 2018; Langenegger, et al., 2019; Miyashita et al., 2021; Yamazaki, Hyodo, & Morita, 2003). Two contigs (TRINITY_m.230427 and TRINITY_m.45429) showed similarity to CRISPs in this study.

3.3.2 Estimation of the mature region of cytolytic linear peptides

As shown above, 15 contigs are identified as a precursor of cytolytic linear peptides (Table 3-1). In general, precursors of cytolytic linear peptides in spider venom are composed of a short signal sequence followed by propeptide and mature regions, which are separated by specific motif, named processing quadruplet motif (PQM). The PQMs in a precursor are recognized by peptidases to produce mature peptides(Kozlov & Grishin, 2007). The sequences in this motif contain an Arg residue at position -1, and Glu residues at positions -2, -3, and/or -4 ($X_1X_2X_3R$, where any of X_n = E). Besides this type of simple precursors, two or more mature peptides are found in one precursor sequence, which are known as a binary or complex precursor(Kozlov, et al., 2006). This precursor structure has been proposed to be a way for preventing potential harmful cytolytic actions during their biosynthesis. In the case of binary and complex precursors, the N-terminal side of mature regions are cleaved by recognition of the PQM, and their C-terminal side are cleaved by an inverted processing quadruplet motif (iPQM), which is a mirrored PQM comprising an Arg residue at position 1, and Glu residues at positions 2, 3, and/or 4 ($RX_1X_2X_3$, where any of X_n = E)(Kuhn-Nentwig, 2021).

Table 3-1. Identified contigs of cytolytic linear peptides

Name	Sequence	Similar peptide	(Accession No, species)
TRINITY_m.2649	MTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMSEMTMKKQESG*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2799	MSANDDEEARIWLTALKFLGKNLGKHFAKQQLAKLGRSEDISENLSADSFQSLEEVKTCLQMTMKKQESG*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2698	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSED MSECLQMTMNKQESG*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2695	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQ LSKLGRSEDMSECLQMTMNKQESG*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2781	MTKKKQESWLTALRIWLTALKFLGKNLGKNLGKHLAKQQLAKLGRSEDISENDDEEARIWLTALKFLGKNLGKHFAKQQL AKLGRSEDISENLSADDDEEARIWLTALKFIGKNLLVNSIKIFR*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2690	MSANDDEEARIWLTALKFLGKNLGKHRISQNSSLQSLEEVKTSLKTCLQSQNNSFQSLEEVKIMSANDDEEARIWLTALK FIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKH*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2667	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRN EQTPEEARIWLTALKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2650	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRN EQTPEEARIWLTALNKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT*	Toxin LyeTx 1	(C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2660	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRN EQTPEEARIWLTALNKFLG*	M-lycotoxin-Ls3b	(P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2706	MKEEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGR NEQTPEEARIWLTALNNASEALGREEIWGFCK*	M-lycotoxin-Ls3b	(P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2678	MKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREEGVSANEDEEARGRLQAFLAKMKEIAAQTLGREDS LSANEDEEARGRLQAFLAKMKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWK AFLAKMKEIASEALGREENLSANEDEEGGFCK*	M-lycotoxin-Ls3b	(P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2715	MSANDDEEARIWLTALKFIGKNLARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEEARIWLTALKFIGKNLG KHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMK SIAKYVAKKQLKKHLGGEN*	M-lycotoxin-Hc2a	(P61508, *Hogna carolinensis*)
TRINITY_m.2734	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSED MSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNE QTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*	M-lycotoxin-Hc2a	(P61508, *Hogna carolinensis*)
TRINITY_m.2795	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQ LSKLGRSEDMSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIY KAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*	M-lycotoxin-Hc2a	(P61508, *Hogna carolinensis*)
TRINITY_m.2773	MKEIASLQMRMKKREIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWKAFLAKMK EIASEALGREENLSANEDEEERMVWLLPLKFLASHVAMEQLSKLGSKIATKLGRNEQIPVISANEDEEERMVWLLPLKFL ASHVAMEQLSKLGSKIATKLGRNEETPVSFFADDDGEERAGLGKIGALIKKAYAIYKAKAAGRNEQTPATVSANDDEEAR KIKWFKAMKSIAKYVAKKQLKKHLGGEN*	M-lycotoxin-Hc2a	(P61508, *Hogna carolinensis*)

In this study, 17 possible mature sequences of cytolytic linear peptides were identified from 15 contigs based on the PQM and iPQM cleavage sites (Table 3-2). For example, the sequence of M-lycotoxin-Lp2 is found in six complex precursors and is repeatedly observed in some precursors. There are four types of PQM sequences and nine types of iPQM in these precursors (Table 3-3). The lengths of linker sequences are ranged from 6-18 amino acid residues. Most of the mature sequences contain Gly residues at the C-terminus, which is likely to be amidated. These cytolytic linear peptides can be divided into six families according to the sequence similarity (Table 3-4).

The sequences of two linear peptides (lyp2370 and lyp1987) identified by mass spectrometric analysis in chapter 2 were found in these deduced peptides (M-lycotoxin-Lp1a and M-lycotoxin-Lp2, respectively, Table 3-4). Although the predicted sequence of M-lycotoxin-Lp2 has the same length with that of lyp1987, M-lycotoxin-Lp1a has the sequence longer than that of lyp2370 (Figure 3-2). This may be due to the additional cleavage of the N-terminal region of M-lycotoxin-Lp1a in the venom.

M-Lycotoxin-Lp1a **MVWLLPLKFLASHVAMEQLSKLGSKIATKL**-NH_2
lyp2370 --------**FLASHVAMEQLSKLGSKIATKL**-NH_2

Figure 3-2. Comparison of the sequences between M-Lycotoxin-Lp1a and Lyp2370.

3.4 Summary

In summary, the transcriptome analysis of *L. poonaensis* venom gland was conducted in this study. This resulted in the identification of 87 contigs, which encode peptides and proteins similar to those in other spider venoms. The number of sequences identified as neurotoxins was the highest, followed by peptidases, and cytolytic linear peptides. As expected from the results of mass spectrometric analysis, the *L. poonaensis* venom contains a large number of linear peptides, which likely show antimicrobial activity.

Table 3-2. Identification of mature peptides found in the contigs

Type	Name	Structure of precursor
Simple	TRINITY_m.2649	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSED>
	TRINITY_m.2799	<EEARIWLTALKFLGKNLGKHFAKQQLAKLGRSED>
Binary	TRINITY_m.2698	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSED>
	TRINITY_m.2695	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSED>
	TRINITY_m.2781	<EEARIWLTALKFLGKNLGKHFAKQQLAKLGRSEDISENLSADDDEEARIWLTALKFIGKNLLVNSIKIFR*
	TRINITY_m.2690	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKH*
Complex	TRINITY_m.2667	<EEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT*
	TRINITY_m.2650	<EEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT*
	TRINITY_m.2660	<EEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNKFLG*
	TRINITY_m.2706	<EEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNNASEALGREEI>
	TRINITY_m.2678	<EEARSKWKAFLAKMKEIASEALGREEGVSANEDEEARGRLQAFLAKMKEIAAQTLGREDSLSANEDEEARGRLQAFLAKMKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREEN>
	TRINITY_m.2715	<EEARIWLTALKFIGKNLARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*
	TRINITY_m.2734	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*
	TRINITY_m.2795	<EEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEDMSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*
	TRINITY_m.2773	<EEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREENLSANEDEEERMVWLLPLKFLASHVAMEQLSKLGSKIATKLGRNEQIPVISANEDEEERMVWLLPLKFLASHVAMEQLSKLGSKIATKLGRNEETPVSFFADDDGEERAGLGKIGALIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN*

PQM iPQM Spacer or linker C-terminal amidation < Signal peptide and propeptide > Truncated C-terminal sequence * End of the sequence

Table 3-3. The PQM, iPQM, and linker sequences in different types of precursors

Type of precursor	PQM	iPQM	Linker sequence
Simple	EEAR	RSED	-
Binary	EEAR	RSED	**RSED**MTVNDD**EEAR** **RSED**ISENLSADDD**EEAR**
Complex	EEAR	REDS	**REDS**LSANED**EEAR**
	EEER	REEG	**REEG**VSANED**EEAR**
	ELER	REEI	**REEN**LFAN**EEER**
	GEER	REEN	**REEN**LSANED**EEER**
		REES	**REES**LSENEN**EEAR**
		RNEE	**REES**VSANED**EEAR**
		RNEQ	**RNEE**TPGSFSADDD**ELER**
		RSEE	**RNEE**TPVSFFADDD**GEER**
			RNEQIPVISANED**EEER**
			RNEQTPATVSANDD**EEAR**
			RNEQTP**EEAR**
			RSEDMSANDD**EEAR**
			RSEDMTVNDD**EEAR**
			RSEEAR

Table 3-4. The structure of linear peptides identified in this study.

Name	Estimated mature sequence	Monoisotopic mass
M-Lycotoxin-Lp1a	MVWLLPLKFLASHVAMEQLSKLGSKIATKL-NH_2	3350.9
M-Lycotoxin-Lp1b	VIWLPALKFLASHIAMEQLSKL-NH_2	2506.4
M-Lycotoxin-Lp2	GRLQAFLAKMKEIAAQTL-NH_2	1987.1
M-Lycotoxin-Lp3a	IWLTALKFIGKNLGKHFAKQQLSKL-NH_2	2880.7
M-Lycotoxin-Lp3b	IWLTALKFLGKNLGKHFAKQQLAKL-NH_2	2864.7
M-Lycotoxin-Lp3c	IWLTALKFIGKNLGKHLAKQQLSKL-NH_2	2846.7
M-Lycotoxin-Lp3d	IWLTALNKFIGKNLGKHFAKQQLSKL-NH_2	2994.7
M-Lycotoxin-Lp3e	IWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKL-NH_2	3851.3
M-Lycotoxin-Lp3f	IWLTALKFLGKNLGLKFIGKNLGKH	2808.6
M-Lycotoxin-Lp3g	IWLTALKFIGKNLLVNSIKIFR	2586.5
M-Lycotoxin-Lp3h	IWLTALKQLSKLGRT	1727.1
M-Lycotoxin-Lp3i	IWLTALNNASEAL-NH_2	1413.7
M-Lycotoxin-Lp3j	IWLTALNKFL-NH_2	1216.7
M-Lycotoxin-Lp4	KIKWFKAMKSIAKYVAKKQLKKHLGGEN	3271.9
M-Lycotoxin-Lp5a	AGLGKIGAFIKKAYAIYKAKAA-NH_2	2251.3
M-Lycotoxin-Lp5b	AGLGKIGALIKKAYAIYKAKAA-NH_2	2217.3
M-Lycotoxin-Lp6	SKWKAFLAKMKEIASEAL-NH_2	2049.1

Table S3-1. List of the sequences of contigs identified by transcriptome analysis of *Lycosa Poonaensis* spider venom gland

Neurotoxin (24)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.59186	MSSGPSLSLSDTAHFRKSCLEIGEVCDGNANDCQCCRSNGFCHCSWIFNHCTCQVGDSSKSYGVCLWKQKNCPNKPGMCTKPCTNRRRCKNRSG	ω-ctenitoxin-Pn3a (P81790, *Phoneutria nigriventer*)
TRINITY_m.1192	MKTSTVFGLCAIVVLMLTIADLSGADEMNSQDAPEERGYCAEKGIRCDDIHCCTGLKCKCNDSGYNCVCRKK	Purotoxin-1 (P86269, *Alopecosa marikovskyi*)
TRINITY_m.38262	MGRWIFAIFLGITLLTQVLLPSAYMATSDADTPAVDDYADVARLLYFARKRSCIKRGSSCDHRPNDCCDNSSCRCNLWGTNCRCQRMGLFQKWGK	U8-agatoxin-Ao1a (Q5Y4U4, *Agelena orientalis*)
TRINITY_m.218293	MKYFLKPLVVTVLLLLCYAVMTNAYVMRDSSLDSYQEPNYEALRQYLLSTRNLNNIPNRRSCIRRGSSCDHRPSDCCFNSSCRCNLWGTNCRCQRAGLFQKWGK	U8-agatoxin-Ao1a (Q5Y4U4, *Agelena orientalis*)
TRINITY_m.9	MSPKVQALLLFVGLITFLVVHAEELSEIEESERACAEEHKSCNWENKPCCDKMSCHCSAAFTNCKCQKGVIRKIIDWIGGK	U16-lycotoxin-Ls1a (B6DD52, *Lycosa singoriensis*)
TRINITY_m.11889	MSPKMQVLLLTIWLFMFLAVHSHEELSESAESERSCANEYQSCDWYNRPCCDNISCVCSWIGTNCECKKGIIRTIRDWINGK	U16-lycotoxin-Ls1a (B6DD52, *Lycosa singoriensis*)
TRINITY_m.16586	MIRYVLISALLVVAVYSFTFEDENEDVFLEEAKEELDPEEERRIALPPGAVCNGHKSDCQCFGAKYKCSCPWLWRFRRSAKCHCKKGWAWTAIKKRSCRNRYQWSG	U2-lycotoxin-Ls1a (B6DD29, *Lycosa singoriensis*)
TRINITY_m.30919	MKILFVLISILYSVHCFTSEEIVGSEHLANELEAVEDINSEQNVALESTREEERSCADMGQDCADDCDCCLNIARCNCWFGKYFCSCTFGNYQTCQAKRGKCTRNRPQSCPKSNINHKKG	U13-lycotoxin-Ls1c (B6DD20, *Lycosa singoriensis*)
TRINITY_m.8162	MKIAVIFGVLLVTLFSYSSAEMFDDFEQADEADELLSLIEEQTRAKECTPRFNDCTNDRHSCCRGELFKDVCTCFYPENGGNEFCTCQQPKHFKYIEKGTDKIKKIGSKIKSWFG	U3-lycotoxin-Ls1a (B6DCQ5, *Lycosa singoriensis*)
TRINITY_m.6715	MSTFTQCKRKVIFNKLFSKQIPLIFPFYRAKACTPRFRDCSHDRHSCCRSSLFKDVCTCFYPENGDKKEVCSCQQPKHLKYMEKATDKIKYLIG	LSTX-D6 (B6DCU5, *Lycosa singoriensis*)
TRINITY_m.6722	MKVFVVVALLVTLISYSSSEGIEDLEADDFLSLMANEQPRAKACTPRFRDCSHDRHSCCRSSLFKDVCTCFYPENGDKKEVCSCQQPKHLKYMEKATDKIKYLIG	LSTX-D6 (B6DCU5, *Lycosa singoriensis*)
TRINITY_m.6716	MNFLKGNINTFRNDCIPKHHECTSNKHGCCRGKIFKYKCQCTTVVDQSGEQAERCFCGTATHHKAVELMAGFGKKLFG	U1-lycotoxin-Ls1ee (B6DCN6, *Lycosa singoriensis*)
TRINITY_m.103490	MANEQPRSDCIPKHHECTSNKHGCCRGKFFKYKCQCITVVDQSGEQAERCFCGTATHHKAVELMAGIGKKLFG	U1-lycotoxin-Ls1e (B6DCJ5, *Lycosa singoriensis*)
TRINITY_m.6709	MKVFVVVALLVTLISYSSSEGIEDLEADDFLSLMANEQPRNDCIPKHHECTSNKHGCCRGKIFKYKCQCTTVVDQSGEQAERCFCGTATHHKAVELMAGFGKKLFG	U1-lycotoxin-Ls1e (B6DCJ5, *Lycosa singoriensis*)
TRINITY_m.6148	MKYQILFGLVFLTLLSYCSSEFEDAGLFMDEEMVEADDPNAIARKDDNENCIAKHHECTSNRHGCCRGKLFKYKCQCVKIVNAQKEETERCACITPGLHKAVEFVLQLFKKAIA	U5-lycotoxin-Ls1a (B6DCV0, *Lycosa singoriensis*)
TRINITY_m.69253	MKVLVLFTVLFLTLFSYSSTEAIHEFDSNAEEDMLSVFACEQVRAKACSPRLHDCSDDRHSCCRGEKYKVICNCYDLEGEDKSEVSTREQPKSEKNIVKLDEKSKTVV	U4-lycotoxin-Ls1a (B6DCT6, *Lycosa singoriensis*)
TRINITY_m.157117	MKVLVIFSVLFITLFSYSSTEAIDEFDSNAEEDMLSVLASEQERAKACTPRLHDCSDERHSWCRGEMFKDVCYCFYPEGEDKTEVCSCQQPKSHKYNEKVVDKTNTLVG	U4-lycotoxin-Ls1a (B6DCT6, *Lycosa singoriensis*)
TRINITY_m.190400	MKLLFFTGLFLLVVVNLIKAEAEAENERAACIRLERQCTKTPGNCCSGLKCECYQRFEKGVAKGKRCWCIEKDVTYKNKK	U6-lycotoxin-Ls1d (B6DCV4, *Lycosa singoriensis*)
TRINITY_m.2157	MKILFFTLSLLLVASFIEVEAKSKGSCLTYAEVCTKNASGCCSDLRCDCYIRYEEGVEKGEKCWCNYKDVIYKKA	U7-lycotoxin-Ls1c (B6DCW9, *Lycosa singoriensis*)
TRINITY_m.2159	MKIVLFAVIAVLLTVSLIEIEAQSKGVCLPYAAVCKTNASRCCSGLTCDCYLRFEEGVQKGEKCWCIHKGVTYTKE	U7-lycotoxin-Ls1c (B6DCW9, *Lycosa singoriensis*)
TRINITY_m.10500	MKLIIFTGLVIFAIANAIEAEAENERACIAEYGECLKAPGNCCSNLACDCYGRFKNGKEIGRNCFCLQKGVIYKREK	U8-lycotoxin-Ls1s (B6DCX9, *Lycosa singoriensis*)
TRINITY_m.10497	MKLIIFTVLVLFSIVSLIKAEAESERACVAQYGECLKEPGNCCSNLVCDCYGRFKNGKEIGRNCFCLEKGIIYKREK	U8-lycotoxin-Ls1s (B6DCX9, *Lycosa singoriensis*)
TRINITY_m.10498	MKLIIFTGLVIFAIANAIEAEAENERACIAEYGECLKAPGNCCSNLACDCYGRFKNGKEIGRNCFCLEKGIIYKREK	U8-lycotoxin-Ls1s (B6DCX9, *Lycosa singoriensis*)
TRINITY_m.3722	MKLGIFLLFSLVVLAVASESIEENINDNLPEQQRACADLNEKCTEGDDCSCCGERGRCDCNWPGKPGCYCMRGGPFDLIAKKFKC	U12-lycotoxin-Ls1a (B6DD13, *Lycosa singoriensis*)

Cytolytic linear peptides (15)

Name	Sequence	Similar peptide (Accession No., Species)
TRINITY_m.2649	MTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMSEMTMKKQESG	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2799	MSANDDEEARIWLTALKFLGKNLGKHFAKQQLAKLGRSEDISENLSADSFQSLEEVKTCLQMTMKKQESG	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2698	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMSECLQMTMNKQESG	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2695	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEDMSECLQMTMNKQESG	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2781	MTKKKQESWLTALRIWLTALKFLGKNLGKNLGKHLAKQQLAKLGRSEDISENDDEEARIWLTALKFLGKNLGKHFAKQQLAKLGRSEDISENLSADDDEEARIWLTALKFIGKNLLVNSIKIFR	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2690	MSANDDEEARIWLTALKFLGKNLGKHRISQNSSLQSLEEVKTSLKTCLQSQNNSFQSLEEVKIMSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKH	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2667	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2650	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNKFIGKNLGKHFAKQQLSKLGRNEQTPEEARIWLTALKQLSKLGRT	LyeTx 1 (C0HJU9, *Lycosa erythrognatha*)
TRINITY_m.2660	MKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNKFLG	M-lycotoxin-Ls3b (P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2706	MKEEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREENLFANEEERVIWLPALKFLASHIAMEQLSKLGRNEQTPEEARIWLTALNNASEALGREEIWGFCK	M-lycotoxin-Ls3b (P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2678	MKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREEGVSANEDEEARGRLQAFLAKMKEIAAQTLGREDSLSANEDEEARGRLQAFLAKMKEIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREENLSANEDEEGGFCK	M-lycotoxin-Ls3b (P0C2U7, *Lycosa erythrognatha*)
TRINITY_m.2715	MSANDDEEARIWLTALKFIGKNLARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN	M-lycotoxin-Hc2a (P61508, *Hogna carolinensis*)
TRINITY_m.2734	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN	M-lycotoxin-Hc2a (P61508, *Hogna carolinensis*)
TRINITY_m.2795	MSANDDEEARIWLTALKFIGKNLGKHFAKQQLSKLGRSEDMTVNDDEEARIWLTALKFLGKNLGLKFIGKNLGKHFAKQQLSKLGRSEDMSANDDEEARIWLTALKFIGKNLGKHLAKQQLSKLGRNEETPGSFSADDDELERAGLGKIGAFIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN	M-lycotoxin-Hc2a (P61508, *Hogna carolinensis*)
TRINITY_m.2773	MKEIASLQMRMKKREIAAQTLGREESLSENENEEARGRLQAFLAKMKEIAAQTLGREESVSANEDEEARSKWKAFLAKMKEIASEALGREENLSANEDEEERMVWLLPLKFLASHVAMEQLSKLGSKIATKLGRNEQIPVISANEDEEERMVWLLPLKFLASHVAMEQLSKLGSKIATKLGRNEETPVSFFADDDGEERAGLGKIGALIKKAYAIYKAKAAGRNEQTPATVSANDDEEARKIKWFKAMKSIAKYVAKKQLKKHLGGEN	M-lycotoxin-Hc2a (P61508, *Hogna carolinensis*)

Protease inhibitor (12)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.201956	MNFQFVLLCMTSVVVVQYAVAFKLIRSKRSTNVPSRCLQDPDSGMCLAYFLNYYYNPSSGECETFVYGGCQGNDNNFWTKDDCEAICKPKTNETEEINSLLLESLESLIEGKDVCEYPPETGECRALFIRYYYDGEKCSPFVYGGCDGNPNNFETEVECMERCGPKSEDESEENEV	U19-barytoxin-Tl1a (W4VSH9, *Trittame loki*)
TRINITY_m.221752	MKNISIFSVVLTLSFSLAVSFTYPRSRRSSDDEDSSSRCQLPPVTGLCRAYFLMYYHNATTNTCNRFVFGGCIGNGNKFWTGEECVEACGGTIDDTMPIQSFHYDYDFEDYDDPCDKPLVVGHCDADIPRFYFDGKECTAFLYGGCGGGNHNNFPTKEKCLEKCGYRILDEENEIENTE	U19-barytoxin-Tl1a (W4VSH9, *Trittame loki*)
TRINITY_m.248055	MEGNMLMVLLVASTIGLALSNPLVSRPNYDNIPQCLQPPASGMCLAYFPSYYYNPSKETCDTFIYGGCQGNANRFFSFDECMAKCGGFKQEPNDSFKLILENGEDQNPSASEQQVEDICSLSTEMGPCRAMMPRFYFNGDKCEEFIYGGCEGNANNFKTIEECQQRCGSTSSSEEVEADNSEETEKASEEAVAEEASVESAEESDSKAEN	U19-barytoxin-Tl1a (W4VSH9, *Trittame loki*)
TRINITY_m.250351	MKNEFLILFLIIGLFCPIYGILESDRLKRLIRIDLNNPLAKDVCDYPLDIGRCFGAKQRYYFNGDTCEEFTYKGCLGNANNFRTKEECMLKCAHKMEGGSGETEEEEHFSFSLFG	kappaPI-theraphotoxin-Hs1a (P0DJ77, *Cyriopagopus schmidti*)
TRINITY_m.41663	MDSKIIAVFLLLVLSTCVLSEKYCPTTRNPYCKKMNIRNNCCKDSDCSSGYCCSEPCGNFCHAPSSIPGGKRVNPNASCKLGYVYR	U14-lycotoxin-Ls1b (B6DD37, *Lycosa singoriensis*)
TRINITY_m.1940	MNSKIFAVLLLLALSACALSEKYCPTPRNTSCKKQNIRNDCCKDSDCTSNAFCCAQPCGNFCIAPSDNPGGRRVDPNASCKLGYVY	U14-lycotoxin-Ls1b (B6DD37, *Lycosa singoriensis*)
TRINITY_m.41662	MPHVLCEFYINKRMVVSNHYIQLSEKELICKMNSKIFAVLFLLALLTCVLSDQYCPKSRLTPCKKMNIRIDCCKDEDCTGGSWCCATPCGNFCKYPINRPGGQRAAGGENCKIGYVYLK	U15-lycotoxin-Ls1d (B6DD42, *Lycosa singoriensis*)
TRINITY_m.11922	MNSHILFVLLVLGIATCVLAGDIELRINICSNELKANRKKCNENCKRLSFQDFGIKKEDRTTLFCFLHQIKPLR	U20-lycotoxin-Ls1c (B6DD62, *Lycosa singoriensis*)
TRINITY_m.240217	MKPTYLFICVISLVYLVSAGRSQFCPVREPIFCIRRQDLCCSHNECSVGELCCEENCGNVCRLAVPRRTSGIEYRYNPNCINESL	U7-agatoxin-Ao1a (Q5Y4V9, *Agelena orientalis*)
TRINITY_m.11914	MNSHILFVLLVLGIATCVLAGGFCPKSRHPHCNLSYKINDCCAQSDCRFGSVCCVEGCGNVCRAESDTPIGEKFVDGSECKLGHVFPKRWYEFW	U20-lycotoxin-Ls1a (B6DD60, L*ycosa singoriensis*)
TRINITY_m.204789	MMFKILLVAIFIVLAESTERKTDCQEHREREAKSKASLPMRLIPECDENGDYKPLQCFKDSDFCACWDKQGNPVTQPSKTIKQCDCLVQKHEVEKKGLRGAFKPSCSEDGRYQKQQCHGSTGFCWCAHPETGEKTSEETRGRPNC	U24-ctenitoxin-Pn1a (P84032, *Phoneutria nigriventer*)
TRINITY_m.217252	MWIPIFLLALAVPTIRSERKCGECVPEKCQPPSEECLAGLVKDLCGCCYVCGRREGELCDGDFLPIPYRNRGYGPCGEHLECRPRTDLAPGDPPEAMCVCLKTETLCGSDGNTYDNECQLTEARYKQREQILKAMHRGPCRSAPKITSPPEEASNYTGGNIAMSCEATGWPIPVFEWRVDIGDGNTIPLPSDDPKVSVQSRGGPSKYEVTSWLQLLSIQPKDDATYWCIAKNDEGESSAAARVVVLDFRGSKTEDNGRINDL	IGFBP-rP1 (G4V4G1, *Cupiennius salei*)

TCTP family (2)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.7640	MIIFKDLITGDEMFTDSSKYKLIDECLYEVECRHVQRRQGDIQLEGANPSQEEADEGTDESVESGLDLVLNQRLVETGFTKNDYKNYLKTYTKSLQDKWKEMGKNDTEIAEAKSKITEAVKKVLPKLADMQFFMGESSNPDGLIALLEYRENPGGDETPIMIFFKHGLDEEKV	Translationally controlled tumor protein homolog (G3LU44, *Loxosceles intermedia*)
TRINITY_m.24068	MRIFKDIITGDEMFTDSSKYKLIDDCLYEVECRHVQRRLGDVQLDGANPSQEEADEGTEEVVESGLDLVLNQRLVETVFSKNDYKGYLKLYTKSLQDKWKEMGKSDSEIADAKSKFTEAVKKILPKIGDLQFFMGESSDPDGLIALLEYRENESGNETPIMMFFKHGLEEEKV	Translationally-controlled tumor protein homolog (G3LU44, *Loxosceles intermedia*)

Non-toxic peptide (2)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.106956	MKCIIALTILATLVVAIQGKFCERSSDCGEGQCCTGSSFNRHCQELSDNGRPCQRPNDYDSYKTGCPCKEGLICSVINYCQEP	U19-ctenitoxin-Pn1a (P83997, *Phoneutria nigriventer*)
TRINITY_m.7591	MRTMKCIISLTIVATLMCLVQGQDQSCGEVYCGEGQCCSGSFYARHCRDYSNDGEPCERPNKYNSYKTACPCKEGMFCNVINRCQKYE	U19-ctenitoxin-Pn1a (P83997, *Phoneutria nigriventer*)

Peptidase (16)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.1011	MDTSVSPCDDFYQYACGRWSKHHELPSDRSYYNTFEVLKDQLKAQLKDLLEEPVTDDDSNATIKAKNLYASCMNEEVIEELKEKPLVLLLDDLGGWPVTSSNWT EENFDWVYLIATLQQYSNNILISQWVGPDGRNSSMHIIQIDQADLGLPSREYYVQGTQQLDAYGRYMVEIAQLLGASPEKAEEEMKEVLEFEIEIANFTIPREERR NYTAMYHKLTLAELQVRIPLVNWTLFFNVAMPLELTDDEEIVIYAITYVERMTDFLVNTPKRVVANYILWRFVYKFVSNLDKRFQAKQQEYYSALYGTQSIPPRWK TCTVYTNKIMGTAVGSLFVKTHFNERSKETAEEMIENIKTAFLELLEEVDWMDTETRSEARQKALLMTEKIGFPDYIMDPQELDKDYDGIEIKPDRYFDNVLLSLKQ YSIREQMKLRTLVDRTQWVSSPAVVNAFYTRPKNFITIPAGILQPPLYHQNYIRSLSYGGIGVVIGHEITHGFDDKGRQFDHRGNLKQWWDSEALIKFQSKADCM IKQYGQYMMTDINMKVNGINTQGENIADNGGVKQSFKAYKSWASRNSEEPLLPGLNLTHDQLFFLNYAQIWCGSMRPEAAINMIRTGVHSPGKYRVIGALSN SKDFSEAYGCPLGSPMNPEKKCHVW	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.15154	MVSKGGCLGWLKKRTNLEKTLLSLSSLLLLTLLLMILIGITYQGKKTKNSPENEEECWSTECIQTAGDILKKIEPEADSCGNFYEYACGTYTKNLGQTISVPKQSTDSV YFTLKRLMESKKNLDVRNLKDVRSFYDTCVKFGSHWQDLKETGDSLNNLMLDLGVGTWPIIDSLYDTSTFSRDKTFASLILLDVPVAFQMEIVPDQQISGSFLVKIS PGSPRESGRSPSDIRGDVHLRSSVLTSFLLLGANAKITDVNDIIAADSSFAAIEQDKNIVCDDVLLSPQEAPSELKRLIPEIDWKIILSEIRNVTGLVRPFAVELHCKQKI RDYVVHLNDLTVRSEANYLGWRFFSKFVEHVVPTFEKSQFGTDDVVPRWRICLELIEKFSPTVLVDLLAKEIINKEVIAQTERISDSSLKAAKRLVSRTQWLTNNQRS RIQEKVSRFRTRYPFKRTAADGDSGKLKIPQVNVNNFTAVIVRQQRKWVIQRFRRLSHEADNVDWIESRVPFLTARSDSVLSDSVLYVYFDKIRTPYIQLKGSQWL NYGGFGVTVAREISRLLLSSSVGQGENTLLWNLWSNSSNTSKCLITSMAKRMGIHEEVLNALADEKLLQNLFLDQSAIEIAYETAKMTAGSSKSLPGLELNDEKMF FLSYAQTQCEADLQVQKSKNIPPIQNRLNSVLQNFDSFNEVFSCASSSHCQIWA	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.19936	MRRADSTLKARDRYEKHSLTKFRNETGICYSPSCISEASNLLKSINMSVDPCSDFYSFACGNWIKNAYLVSTLESQFTKLTRIIHLRVKSILEDLEKEDNLPDHKMKTS DFYKACMDRDHRNEIGATEFLNDLKKFGGWPLIEEEWDAANFNWSNVMIQLFKNGYISNMLFTVKLEIDAKNTSRSIIAVMQPSLGLKNKLYYIIPDGQIATSYK KAIFEVAKHLHPNANLVKKTSMDAIEKAFELESKIADLIKEVNMRYPDQFYNIMTLEELENFAPKIPWRNLLKHLLPPETILTNEERILVTVPSMIKGLNKILEEMSTE QGKRILSNYMFYRTAFFAASNLNTEFYNLIKGIFNTDSVANNQLWEICVKDTMDLFPLPLTSSYVAKYLKNNTKELLEVMFDDIKNALVEDIKNAAWLDTETKNYA LFKVKKMKAAIAYKEEIMKDENLNSFYKEVNVTDNHFSNVKQLIAARQQRNMQGLHKKRDRFRWTDVNNLLTANAYYQPSSNAIIIPVGVLHDTIFAAERPNYI NYGSLGMIVGHELTHAFDSIGSLYDADGSYRMWWKKESWKKFEERVKCYKEQYDKFYLSEADSNVNGTLTLGENIADNGGLQTAYYAYNLLLKRLGSEKSLPGL PYTERQMFWISFANVWCRKQTEESIQYGMIYSVHSPSRFRIIGSLSNMKEFSEDFNCKLNSTMNPKSKCLLWK	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.9580	MAGNLAVAFRNVTAQKCAWWNRRSPCEKFLFILCILLLSTCAVLIIVNLITEKRLRVAEDIVSGICHSEECIKAAARILERLDTSTKPCDDFYHFACGNFLHRHTVPDD HYLRSTMQTMQDDLYVTLKRLIEHPQSSNDSEAIIKVKKLYTSCMNTSSIEDGSVKVLQDLLTNWGIGEWPILIPRWNKTVDMEWRLAMLHIHEVKPFFSTFVAP DDRNSSVYLLHINAGGPFLNPQYFLNTSDPDYVRYVLSYKNLIFETVRLLGVQESDAKKDIEEMLEFEVEFANISQDDPFDSINETSQLDDDFVFNKINVSVLEDMIP EIKWNVLIEYIFDYIGLSTEKIDISFVVHCEKYLRHLVALLNRTPPKTAINYLTWRFVMKYMPYLDIHFRRLYYDFRREVPVRAEERTFFARWKECVHIATEGFGMAL ASLYVKDEFNEEMENEIKFLISSLKEAFIKSIHKQTWLDHDTKVLCEEKVSAMGNKIGFPRYILDPFQLDADYVGLDVTEDHFLDNILKINRYEETKDLSKITRTVDKE RDWFVQPLVVNAYYEGSGNTVIFPVGILRRPIFTPRRPKYLNYGMLGVIIGHEITHGFDNNGRKYDKAGNFTQWWTDEIVEKFKEQVSCYIEQYSQLPIDMVGQ NVNGNQTLDDNICDNSGLAQAYRAYRHYVSKYGEEPRLPGISYSNTQVFFLQYAQLWCEILSKEANERFIKDSHSPGKYRANIPLMNSPEFSSAFNCPTGSPMNP VKKCRLWA	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.19952	MFTSQYMNFDEKSRLYPERGPGKTFPVTTPLERTLLIIIAIVVLIGIGLSIGIVSLKKARDRYEKHSLTKFRNETGICYSPSCISEASNLLKSINMSVDPCSDFYSFACGN WIKNAYLVSTLESQFTKLTRIIHLRVKSILEDLEKEDNLPDHKMKTSDFYKACMDRDHRNEIGATEFLNDLKKFGGWPLIEEEWDAANFNWSNVMIQLFKNGYIS NMLFTVKLEIDAKNTSRSIIAVMQPSLGLKNKLYYIIPDGQIATSYKKAIFEVAKHLHPNANLVKKTSMDAIEKAFELESKIADLIKEVNMRYPDQFYNIMTLEELENF APKIPWRNLLKHLLPPETILTNEERILVTVPSMIKGLNKILEEMSTEQGKRILSNYMFYRTAFFAASNLNTEFYNLIKGIFNTDSVANNQLWEICVKDTMDLFPLPLTS SYVAKYLKNNTKELLEVMFDDIKNALVEDIKNAAWLDTETKNYALFKVKKMKAAIAYKEEIMKDENLNSFYKEVNVTDNHFSNVKQLIAARQQRNMQGLHKKR DRFRWTDVNNLLTANAYYQPSSNAIIIPVGVLHDTIFAAERPNYINYGSLGMIVGHELTHAFDSIGSLYDADGSYRMWWKKESWKKFEERVKCYKEQYDKFYLSE ADSNVNGTLTLGENIADNGGLQTAYYAYNLLLKRLGSEKSLPGLPYTERQMFWISFANVWCRKQTEESIQYGMIYSVHSPSRFRIIGSLSNMKEFSEDFNCKLNST MNPKSKCLLWK	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.60402	MHNNKGGDEEKTDPKENDKCMENNEHDKDSEVNQKSSKLDALRGRAKQLASNKYHLIIVLLCVLLFILFLIVIILAVQLGTYACEVPNECRTADCLRAASFVLDRSE PSEDACDDFWAYSCKNWISQNPVPANKGSYSVTDELQDKIYNRIRHLIDLIRHDVDSSSIEMKVKTFYDSCRNMHNIERRFPDDIKRAIYDIGGWSLIEYARPLQW DRMEVLSDLHSVYGIPVFFQVVVEPDDLDPQRNIIKLIPGGLGLPNRDYYFRPHDDTYVKAYKTFMIDTVKEMGVGLSKANQFADEVFNYEKRIAEITPSTEELRS NPNLYVVKTIKDLDIPPSVIQWTKFLQAYFKDSADPETEIGVLSENYFLDISRIISSTADSVLNNYYMWRFLHTFVPFSSSRFRLVANYFKQTFEGVPDLNLKYKENW ESCIDETVKYLGHAVGAMYVSHYFPRESQDEIQLYLSKLAGAFVSVSNNMPWLEEEHSSALSKVKSLLLLSGHPAFIKNNTLNSYYKELRVSIDFLKNIESGVHFLHK KQEEMLKSRATTDYSWTIYSHDVMAHYKYAGNQLVVPAGLFESPLFDSKSPLSVRYGVLAAHVANKMAEAFDDKGINYDEFGAIKNWLNNDTHLSFASRKKCL KSVISNSFIDDIQANADLTIGSFIADIGGIKIAYEAYKRHAEEDQEFARLPNVGVSNKQAFFISYAQSLCQSIRPEKLKSYRDANVELPAELRVLSTLQQLPEFSEAFSCS RDRSMNSQEHCRMW	Neprilysin-1 (W4VS99, *Trittame loki*)

TRINITY_m.28887	MAVTVTQSQLTLTSGRDRFRNPPWWNRRSRLERILCGVAGICLLMCTAMAVALAVVGYHYQVMKTGNGVNEAGLPSKTDKLVYSLYNAIYPPKEIGICLTPGCV KAAATILNNINEKVEPCENFYEFACGGWLQKQLIPDDRSSVSVFSLLQDDLDQLLRSMVETGIKDTDPQYIQNLKSMYESCMNTTHIEMMGNEPLLRVVKELGG WPAVEGDKWNSTNFDWMEALFALRKVGFGHNIFLSVSIGPDIRNNTRHIVDLDQASLGMPDRNYLLKGVEDPLVSAYYQLMVDSAVLLGANKTKAQEEMKSA LLFEIAIANFSVPREERRNISRLYNKMTVDDLYKLAPDLNWDKYFNKLLMDKISRSEEVIVIVPDYVQNLEKLMLSTDKRVIANYLMWRVVGQAFPTLHRAWGEIS QHYSSILTGKVRQEARWEHCLGTLSGSLSTALASLYVRHHFQDGSKDLALEMVNYIHREFLHVLSDVDWMDSQTKERAKEKAQNMATYIGYPNELLEEWKVSEI YDGLHLNSSSYFENIRILRKWATDYVLSKLRKPNVKGDWKKRSAAAVVNAFYNSIENSIEFPAGILQGVFFSKDRPNYLNFGAIGFVIGHEITHGFDDRGRQFDKD GNNINWWEPETDSTFRERAQCIIDQYGNYSVDEVGLQVNGINTQGENIADNGGIKEAFRAYLQWIRDHGAEDYLPGIKYSPTQLFWISAANVWCGKYRPEVLK LRIMTGSHSPAPYRVIGPLANTPEFAAEFGCPLGSPMNPVKKCTVW	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.43613	MFIDDEMPRYKRTDFEEEETSSMASMAVDAPMNPPSSQGVQVRYSPGRRRNIWERASPLEKILVVCVLVLVLLVIILSAVLNSYSSSGTVKVVHITAKNGSSSGN NATEYCVTPACVTVAAAILNAMDQTVDPCTDFYQYSCGGWIRSNPLPDDKSIWGTFGKLWQENQIVMKNVLEDDSFKLTSEAERKARIYYMSCLDKNETVDKL GAKPIIDLLSKIGGWNISGDFNISQWNFQRTLELLHNQYNRGGLFSWGVGEDERNSSRNILQLDQGGLGLPTRDYYLNKSKDDEVLVAYLNYMTKVGVLLGGEE NATRAQMEDVIEFETKLANITIPADERRDDEELYHKKTLNELRDLAPVLDWASYFNSAFRRINREITPAQELVVYAPEFIHKMSELVTQYLSSNEGKVVISNYLGWS LVQSLTSCLSKPFREASKILRKALVGSEGGESPWRYCVSDTNEVIGFALGAMFVREVFHGDSKPMAENMINEIRDAFKENLPLLKWMDPETRKLAKEKADAITDM IGFPEFILDPKQLDKKYEGMDFSEDEYFNNNIEVGQFALLSNMRKLDKPTNRSEWKMTPPTVNAYYTPTKNQIVFPAGILQAPFYDPNYPKSLNFGAMGVVMG HELTHAFDDQGREYDKYGNLHQWWKNSTIESFRERTQCIVDQYSSYEVNGESLNGKQVLGENIADNGGLKAAFHAYQDWVKSHPIELPLPAVPLSNNQLFFIGF AQVWCSTSTPEAMHLQILNDPHSPAKFRVVGTLSNSVDFAREFKCPPKSAMNPENKCEVW	Neprilysin-1 (W4VS99, *Trittame loki*)
TRINITY_m.252247	MYDYDIALLQMERPAEYNEYVRPICLPPKELPDDSPLYKPKSSAWATGWGHDAKISRNDRVIPRTTENLKELYLPIQSKERCIESTRSNGVNISFFTDRMFCAGDGS GGNDTCQGDSGGPLMQAQTNAEGYIYWTQVGIVSWGIQCGMPNTYGYYTYLQRFVPWVEETIENALGAVGM	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.77294	MIFHVSLVTAISLLFGIKYPVFMTHGVFNISHDLALARLNAPVLMNPHVQVVCLPQLNEELQAGERCFATGWGATRGSKSENVLKQAEHPIQDETKCYNEYQKFE SETMICSGRLEPLHGLCHGDSGGPLVCKKDGKWHVYGVASFITDTNFISGLCGLANKPTVFNKISAKVEWIKSIISNNN	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.183905	MSSGEIVCSGTIINKRFVMTSAACVTTSKSIPKKPNEISVVLGLSNRNGGDQSQRFGVKAIYVHPGYRTLNYDNDIALLELEKPMVYSRYIRPICLPTDSSNLYKPQSP GLGLGWGSTSPWGNAGTLKRNSLKEIKVTIQHPDACKEYLGEKYARYFTNLVFCAAEESPALCSGDSGGPVMQSEMREDGHRIWRQVGIVSFGMGCGPNNS MGFYTYVYKFLPWIQDIADNSN	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.183910	MAAVFYVKDSVTKDDIVCSGTIINKRFVMTSAACVTTSKSIPKKPNEISVVLGLSNRNGGDQSQRFGVKAIYVHPGYRTLNYDNDIALLELEKPMVYSRYIRPICLPT DSSNLYKPQSPGLGLGWGSTSPWGNAGTLKRNSLKEIKVTIQHPDACKEYLGEKYARYFTNLVFCAAEESPALCSGDSGGPVMQSEMREDGHRIWRQVGIVSF GMGCGPNNSMGFYTYVYKFLPWIQDIADNSN	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.61171	MFTAICILWSLQLIQGSPLDLSLEDTRPLCPCGDYASSVNTRVVGGHAVPKHEFPYASNLLFRPFEEYSSRQASQKLSPFCGGTLITDRHILTAAHCLKDRDPEDLAV DVGDYSLRELKDRQVVNVRNVTKFPEYRKDSFHTDIGIIELEHPVRFRQGMRTALLPNEGLKLKPGTTVSVYGWGRLSYYGGHPDVLNTVDLPVVRNEECQHEFI SRIEPSMICAGGQEGKDACIGDSGSGLIVRLDNEFVLCGVVSFGRKCALPHVPGVYTRVSSYTDWILDQTRSASCRPCVYDE	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.71477	MFADSKRPRFCMLHRNTTLNSYFLLSILFTLGCVAKGGRPKTVKNCGIVDVPQARIIGGIWAQHGDYPWMVSVHEYYRGFDHVCGGTIINEHWILTAAHCIDYP TKPRKYEIYVGLHRLSKKDRKVSKKHKISKVIVHEEYDKETFLNDIALLRTKEPINFRASRGYVNGICLPVTDKDPKGWAIVTGWGHTLEGGDNSDVLKEVKVPIVPR ELCNEAYDDDPDDDIEENIFDTQLCAGAPNRDSCQNDSGGPLIQRSKNGVHTLIGVVSYGSGCGDRHFPGVYTKVASYMDWMYKNMKD	Venom peptide isomerase (Q9TXD8, *Agelenopsis aperta*)
TRINITY_m.4971	MQLVFAAFLLLLTGYTSGKLSTVKDCGKATVPQARIVDGEVSKRGKYPWMVSVQQWDGDKLRPICGGAILNENWIVTAAHCFDQPVKNSDYEVYVGLFSITKT NEPPVRKHKVSKIIIHEDYKDTGFANDIALIKTATPIDIKGSKGYVNAICMPSGITNPAGDAIVIGWGTIYDDGPLSAELREAVVPIVPWKKCKEIYGNINSDFEYVQV TPFVMCAGGTGKDSCQGDSGGPLFQFDKNGVATLLGTVANGGDCGNGRYPGMYMKASAYKSWMDRVTT	putative PQM protease precursor (AUI10826.1, Geolycosa vultuosa)
TRINITY_m.4960	MQLVFAAFLLLLTGYTSGKLSTVKDCGKATVPQARIVDGEVSKRGKYPWMVSVQQWDGDKLRPICGGAILNENWIVTAAHCFDQPVKNSDYEVYVGLFSITKT NEPPVRKHKVSKIIIHEDYKDTGFANDIALIKTATPIDIKGSKGYVNAICMPSGITNPAGDAIVIGWGKIYDGGPISAELREAVVPIVPWKKCKEIYGNINSDFEYVQV TPFVMCAGGNGKDSCQYDSGGPLFQFDKNGVATLLGTVANGGDCGNGRYPGMYMKASAYKSWMDRVTT	putative PQM protease precursor (AUI10826.1, Geolycosa vultuosa)

Metalloprotease (4)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.136344	MKVINWNIFVMNLLAEDPMVTEGFFEGDIDTENSFQMLNGISDDNLRWPNAIVYYAIEENLHHIKPLIQKAMQHIESKTCIGFVERTNEADYVWIHIGKGCSAQ WGRRGGKQRLSLGHTGCHTLGVVVHELMHAIGFQHEHSRSDRDDYITIFLENVRPGQEHNFEKLEPSENRLLTQFDTSSILLYGSTTFARPNTYSMLLKDGSVMV EVYKKNSMTNLDAEKVNKLYQCT	Astacin-like metalloprotease toxin (A0FKN6, *Loxosceles intermedia*)
TRINITY_m.44881	MVAALVFILLSIGCSLVTSKAYLGDLPIQNPDLFGGDILGIEDAEDRNAIVDKRQIWPGGVVPYEQDPGLKATVFRLTLEGAFDQYKKSTCIKFVPRTNEKDYIRLFP GEGCYSHVGRTGGQQPVSLGQGCGWMGTIVHELGHAIGFYHEQNRSDRDDWLIIYWENVKEGMEDQFFKLKPDQNQLLTPFDYDSIMLYGSYTFSKDRKKLK TMVGKNNEFLQEVISKYRLSKSDIQRINTLYNCKM	Astacin-like metalloprotease toxin (A0FKN6, *Loxosceles intermedia*)
TRINITY_m.35058	MAPFFYFTCVVLWLGSSSATEDFLGDLPMQNPDLFQGDILGVEDPEDRNAIVDKRQIWPGGIVPYVLDPGITKTVSPAILTNAMWLYKRDTCIRFVPRTNEANYI RIFPGQGCYSHVGLTHVGAQPVSLGQGCGYMGTVIHELAHALGFYHEQNRSDRDDWLTIYWDNIKEGMADQFMKLKPDQNQLLTPFDYDSIMLYGSYTFSKE YGKLKTMEGKGGMFLKDVIRKYFMSKSDIQRIKALYKCPS	Astacin-like metalloprotease toxin (A0FKN6, *Loxosceles intermedia*)
TRINITY_m.220610	MFALAALSLFLVITTGLSRPEKDGPNVHPEYLDHLLDTEQYELFAKVMNIPDASLVRPKEGERSENPMFNKDLEMGDIVPAFGPLKEESRQGINFDRYPGSKWQN GELHFYIKLNDFTESQLQTIYDAIDSWNRVVRCLRLIDLRETVRKISDYVYVFGGNGCWSYLGRIGGEQGLSLQPNGCVWKGTVLHEFNHAAGFAHEQNRLDRD DFIEMLWGNIPDDWKSQYEKTNPRDYGLQCSYDYYSIMHYGIKAPGTNKDAFRVKQDGIDQNKIGRSTEFTKIDIEKINKLYC	Astacin-like metalloprotease toxin (A0FKN6, *Loxosceles intermedia*)

Acetylcholinesterase (5)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.101805	MRRDTLLFCLLYLSYLLPCRSLLSRRVVTTPSGRVEGTVQSHNGVSVNAFLGIPYARPPVGDLRFKAPRTPAPWEGTLSAKEMRSGCVQFTRNPFPWRDVPPGL SEDCLYLNVWAPSTLLNTINRRTVMFWVYGGAFLFGSSRVDFYDGHVLAVEGDVVVVTVNYRLGPFGFLYAGNEDAPGNMGLLDLVEALKWVRGNIASFGG DPNSITLFGQSAGGIAIGMFCVSPLTRGLFKRVIMQSGSPTNLGGEDNNLDAKMSQRLAEAVGCAEDENSLADDPETVVECLRNVDAFELARTFSALGGNGSR LVYPRFGDELLPQNPRKAIAKGDFGAVDVLIGSNSGDGALIITDSMTEVFGFFGEKDPQINRTFGTSIIKRQFQDFVAPCSVVKQYLGGVQDDDFDGIRQQVFM ATGDYERTCPATYFAESFAKKGVKAYYYFFTHRPSKSPFAPWMLASHFDETPFVFGYPLRYTRNYTIRERLLSRRMIQAWTAFAKTGTPIGGVWPQYSRENPRY QILDVNSLGTGMGPRQNNCDFFRPFFGF	
TRINITY_m.126349	MRLFLLLSIITLASCDRTVTTTNGPVKGISLSPKKVVAFLGIPYAEPPVGPLRFAKPIPKSFSEEVYLANEFPPSCVQASLGDFYFAPDPKGKMSEDCLYLNVWVPEG GSDTKPIMLFIHGGAFMVGSSNMKVYDGANLSEHGDVIVAAINYRLGTLGFFSGFVEDAAGNMAMYDQVMAIQWIKNNAKHFGGDPDNIVLFGQSAGGFS TSLHLISPLSKGLFKRAILQSGTVVSPLFKSNDVLQMTSQYLASFLGCTDENNSLNKNPQSVIDCMKALPTEKIIQADGILLKSGGMMFPRIGDSFLPENSVDSYK NGDFKDTEILLGVTSNEGSLFLVFHMTDTLGSFGQKIEVQDFNETSAMQYFKRISQHNSKGMPSEIVQEYFDRVKKDPGYTYLQATSDLMGDIIITCGAVFQAD YYSLRDNPVYFYVFDYRPDSTPLAEWMGVAHFDEMQYVFGNPLHQEFTEYEKEFSGDIMDMWVAFAKTGNPNIPGGVKWPRYSYQDPKYLVLSKRDRVAV KPDNNSCEKWRKVFETEIKEDTLKKLKKPA	
TRINITY_m.229987	MRLVCLLLLFGCTLAFDYEGYGCNDNPYIRDAGPLVRTSSGSLFGKYVKVFDTKVAAFLGIPYASPPVSEKRFLKPTPLERWDGVLDATVKPPPCMQYSSRNFSW IPKSTPSEDCLYLNVWTPAKCCCTGSERLPVMVWIYGGGFYSGSTDMDVYDGSTMVSYSKAIVVSINYRVGIFGFLSSGSDEAPGNVGLLDQNLALKWIKKNIG YFGGDPERITIMGESAGAMSVSLHMVSPMSRNLFQAAILQSATAYQPQMVDNKAMSYIKANTFAKVIGCLQGNSSIESDPEVMECLKAKSSWELARAEDVFT SKLPQYFLPTYGDEFLPTSPFKAFNEGHLAPVDILIGNNRDEVTIFLNYMVPDIFRLDSEPQLNLEAAKQIIHGLFQTVPRPSATEIFDHYFQNVSEGDSSEIVKTIA HAFGDYVFYCPVALFAEKARNAYVYWVSHRSIRDQNPEWLGVPHFYEVPFVFGVPIQDTFTYTPEDGRFSADLIKKWTTFAEHGHPLGDRNPDDWPPYSAEN PMAYELNPRQRGLLSFPRADACEFWRPYIVPKTC	
TRINITY_m.130212	MKFLFIIQSCFTVVTFRLLFAPVFGEESNFGQQNSRYFAEPEFKYYTTQNFRYSYTEFDHAAHDERCGPIYESKPTHKITPLGIVRGRQVFLCDRPGIPLRQRPGQI GKYIHSVQVFLGIPYAEPPVRRRNEENLQFKEPLEKQKFIDIDASSYKPACPQHKKYFSKGQGINTTSEDCLYLNIFSPVQEDSDSLLLNGYPVMIYIHDGNWDHG SGALFPGHMLAASQKVVVVTFNYRLGPLGFLATGDSNSPGNYGMLDQVAAINWVYQYIKAFDGDPDKITVFGSGVGGASAALLALSRLTKRKIKRVIAQNGSP VADWAAISDPRFMQNTTEIVGEGLGCSQQDTAHLVECVGTRLNEIKILEIKPRIGWLPWGPVVDNFTRRSLDQFLHESPEDMLEKGIDFRKDFAFMTGVNRD QGASIFLEDVITYRDLDITSQVFDKKIRSFMKMYNYSVNAEGIFDAIKFMYKPPLDSYNETLLREAYIHLLTDRYYVAPMEKIVHLMVKNKVPTYAYVLNYTLQGY STILKDFISPEVDYLLLSGAPFMDPKYYPAYLNLKNAKWTEADRNMSQAMMEAWANFARDGEPTKTRLFNTIMWEKVEENYLRYLNFNATNTSSEMIYKYRE RESRFWNFFLPFFIDREPPTLAPTLEPGVAELRLLTSALWGSVFVAVLLIVITLVCCSLYCRVRRYDSY	
TRINITY_m.130210	MKFLFIIQSCFTVVTFRLLFAPVFGEESNFGQQNSRYFAEPEFKYYTTQNFRYSYTEFDHAAHDERCGPIYESKPTHKITPLGIVRGRQVFLCDRPGIPLRQRPGQI GKYIHSVQVFLGIPYAEPPVRRRNEENLQFKEPLEKQKFIDIDASSYKPACPQHKKYFSKGQGINTTSEDCLYLNIFSPVQEDSDSLLLNGYPVMIYIHDGNWDHG SGALFPGHMLAASQKVVVVTFNYRLGPLGFLATGDSNSPGNYGMLDQVAAINWVYQYIKAFDGDPDKITVFGSGVGGASAALLALSRLTKRKIKRVIAQNGSP VADWAAISDPRFMQNTTEIVGEGLGCSQQDTAHLVECVGTRLNEIKILEIKPRIGWLPWGPVVDNFTRRSLDQFLHESPEDMLEKGIDFRKDFAFMTGVNRD QGASIFLEDVITYRDLDITSQVFDKKIRSFMKMYNYSVNAEGIFDAIKFMYKPPLDSYNETLLREAYIHLLTDRYYVAPMEKIVHLMVKNKVPTYAYVLNYTLQGY STILKDFISPEVDYLLLSGAPFMDPKYYPAYLNLKNAKWTEADRNMSQAMMEAWANFARDGEPTKTRLFNTIMWEKVEENYLRYLNFNATNTSSEMIYKYRE RESRFWNFFLPFFIDREPPTLAPTLEPGVAELRLLTSALWGSVFVAVLLIVITLVCCSLYCRVRSYKRPPDDLDSSREVIVNHSATQQDTAV	

Cysteine-rich secretory protein (2)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.230427	MHIFSLFLSITILVTYAVAQTNTECPELFLRYSSTHSYCKTDNPKCNILRSFINEEEKNTIVRVHNSYRSRLAIGEEQNLPTAGDMLQMVWDDELAAVAQKHAQQCL FEHDCANCRRVQNFGVGQNLAIEQVFNSPSVPNSNWPSAIKSWYDEIAYFSSNAIDPFIAPPPEPTYGHFSQMFWASSWRIGCGYVLYQDRNMYQKLYVCNY GPAGNTVSQSMYKVGSPCSACPLNSCCDKTCGAKAEYSGLCKITDNKAPKYKPVVPYIFYCDFNNQSDCQSYVTGVNKWLVYNTLSGSFLSIILKGGEESSMTFQ NKIKPSTSAFCFVVMYRKGPNEDGKEEASTATETFSGSGFEFQQNLPTFNGDVKQQFSQFSMTLSWNMETEIKITFAVPAGKPDQFFEIQSIYATEGLCKQ	CRISP/Allergen/PR-1 (W4VS53, *Trittame loki*)
TRINITY_m.45429	MSAPLGIPSMFLALCALLCVLVAVQSAYCPDPYKRYSKDHSYCVGVCCSCAKYRRGVSPWLKQELVRLHNNLRSKVAGGKSYGVDHLPRATNMLEMVWDDEL AEVAQRWSDKCERKSDCEDCRRVGRFGVGQIIFDFDGKMEGEDIENEFFSRFQGLQSFRKEQVERFTAGTNPKTAQMLWAKTWRIGCGFLDFNYPGNGRRYT QIVCNYGPNGNVEGEEVYKGGSVCSACPANTCCGDSCKKYNLKANYHGLCKVIDENLPPEGNVPHKKTGNEVFYCGFNDESDCSHSVEGVDRWIRNVTTGGT WLNTYLGNYGHTVLKFNQPIISKSGKLCLAIKTRSGPMEAGQPYRYQLSGGLEEEGKFSTSFVFPPADDNVKHKFHTDYFQPSEFPKNRGAKLTLSFQVSGEGPKT YLEIDSIVALEKDCPKKTITEGTGFITSP	Venom allergen 5 (A9QQ26, *Lycosa singoriensis*)

Techylectin-like protein (5)

Name	Sequence	Similar peptide (Accession No., species)
TRINITY_m.72849	MSFSSNISDEAQASIVVTLLNNARNLLDDAKKYATMKTKLSAAENVYYRPMDCAEVREKNPLARSGVHKIWPHSRAVNGSVEVYCDMEDDGGSWTVIQRRGD FDSESDYFFKEWQFYKVGFGDRMKDFWLGNDVIFALTNQRLN	Techylectin-like protein (P85031, *Phoneutria nigriventer*)
TRINITY_m.257368	MIFALTNQRLNEVKFTLTDWESNTTYATYDEFWIDDEDHKYTAHVSGYKGTAGDSFYNTNGMPFTTKDQDNDKYEKNCAVQSRGAWWYGRCHMSNLNGM YLKGFHESNADGVNWYSLRGHHYSFKDTVIKIRPKGFQKNQCGTTSL	Techylectin-like protein (P85031, *Phoneutria nigriventer*)
TRINITY_m.72846	MLSRLLVWVNLMGVMYLGIYADDPQFISPLSRNITDQTQSSIVVTLLNNARNLLEDAKKYASIKTRCYAAENVYGRPMDCAEVKEKNPLANSGVHKIWLHSRAV NDAVEVYCDMEDDGGSWTVIQRRGDFDSGKDYFYKDWQFYKKGFGDRMKDFWLGNDVIFALTNQRLN	Techylectin-like protein (P85031, *Phoneutria nigriventer*)
TRINITY_m.72851	MLSRLLVWVNLMGVMYLGIYADDPQFISPLSRNITDQTQSSIVVTLLNNARNLLEDAKKYASIKTRCYAAENVYGRPMDCAEVKEKNPLANSGVHKIWLHSRAV NDAVEVYCDMEDDGGSWTVIQRRGDFDSGKDYFYKDWQFYKKGFGDRMKDFWLGDSFSAVNGLPFTTKDQDNDKYEKNCAVQYTGAWWYGSCHLSNLN GKYLKRVHKSHADSVNWHGFRGHNYSLKDTVIKIRPKDFQKNLSGITPL	Techylectin-like protein (P85031, *Phoneutria nigriventer*)
TRINITY_m.265263	MINHWVPVMGILLFCTDCGAADSTCGEKVSLYYHDAAADLIAKAKYHFPSRLHCENLSRNAVDCSEWLEKGYKKNGVYTVWPQSRVVDGKPLEVYCDIETDGG GWTLIQRRGSFGRAVDYFYKDWKSYKEGFGDTEKEFWLGNDNIFALTNQGLYSARFDMKDIDNETRYAHYDKFWIDDEYNKYTLHIADYRGDAGDSLVEHHN HQKFSAKGEDNDNQKDENCVEKYRGGWWYNSCHESNLHGLNLKGKHESFANGINWKSWKGYYDSLQSTEIKIRPRNFKLPHSSLNNTPM	Techylectin-like protein (P85031, *Phoneutria nigriventer*)

References

Adao, R., Seixas, R., Gomes, P., Pessoa, J. C., & Bastos, M. (2008). Membrane structure and interactions of a short Lycotoxin I anaiogue. *Journal of Peptide Science, 14*(4), 528-534. doi: 10.1002/psc.993

Cassola, A. C., Jaffe, H., Fales, H. M., Afeche, S. C., Magnoli, F., & Cipolla-Neto, J. (1998). omega-Phonetoxin-IIA: a calcium channel blocker from the spider Phoneutria nigriventer. *Pflugers Archiv-European Journal of Physiology, 436*(4), 545-552. doi: DOI 10.1007/s004240050670

Grishin, E. V., Savchenko, G. A., Vassilevski, A. A., Korolkova, Y. V., Boychuk, Y. A., Viatchenko-Karpinski, V. Y., . . . Krishtal, O. O. (2010). Novel Peptide from Spider Venom Inhibits P2X3 Receptors and Inflammatory Pain. *Annals of Neurology, 67*(5), 680-683. doi: 10.1002/ana.21949

Kabanova, N. V., Vassilevski, A. A., Rogachevskaja, O. A., Bystrova, M. F., Korolkova, Y. V., Pluzhnikov, K. A., . . . Kolesnikov, S. S. (2012). Modulation of P2X3 receptors by spider toxins. *Biochimica Et Biophysica Acta-Biomembranes, 1818*(11), 2868-2875. doi: 10.1016/j.bbamem.2012.07.016

Kozlov, S. A., & Grishin, E. V. (2007). The universal algorithm of maturation for secretory and excretory protein precursors. *Toxicon, 49*(5), 721-726. doi: 10.1016/j.toxicon.2006.11.007

Kozlov, S. A., Vassilevski, A. A., Feofanov, A. V., Surovoy, A. Y., Karpunin, D. V., & Grishin, E. V. (2006). Latarcins, antimicrobial and cytolytic peptides from the venom of the spider Lachesana tarabaevi (Zodariidae) that exemplify biomolecular diversity. *Journal of Biological Chemistry, 281*(30), 20983-20992. doi: 10.1074/jbc.M602168200

Kuhn-Nentwig, L. (2021). Complex precursor structures of cytolytic cupiennins identified in spider venom gland transcriptomes. *Scientific Reports, 11*(1). doi: ARTN 400910.1038/s41598-021-83624-z

Kuhn-Nentwig, L., Langenegger, N., Heller, M., Koua, D., & Nentwig, W. (2019). The Dual Prey-Inactivation Strategy of Spiders-In-Depth Venomic Analysis of Cupiennius salei. *Toxins, 11*(3). doi: ARTN 16710.3390/toxins11030167

Kuhn-Nentwig, L., Muller, J., Schaller, J., Walz, A., Dathe, M., & Nentwig, W. (2002). Cupiennin 1, a new family of highly basic antimicrobial peptides in the venom of the spider Cupiennius salei (Ctenidae). *Journal of Biological Chemistry, 277*(13), 11208-11216. doi: 10.1074/jbc.M111099200

Kuhn-Nentwig, L., Schaller, J., & Nentwig, W. (2004). Biochemistry, toxicology and ecology of the venom of the spider Cupiennius salei (Ctenidae). *Toxicon, 43*(5), 543-553. doi: 10.1016/j.toxicon.2004.02.009

Kuhn-Nentwig, L., Stocklin, R., & Nentwig, W. (2011). Venom Composition and Strategies in Spiders: Is Everything Possible? *Advances in Insect Physiology, Vol 40: Spider Physiology and Behaviour - Physiology, 40*, 1-86. doi: 10.1016/B978-0-12-387668-3.00001-5

Langenegger, N., Koua, D., Schurch, S., Heller, M., Nentwig, W., & Kuhn-Nentwig, L. (2018). Identification of a precursor processing protease from the spider Cupiennius salei essential for venom neurotoxin maturation. *Journal of Biological Chemistry, 293*(6), 2079-2090. doi: 10.1074/jbc.M117.810911

Langenegger, N., Nentwig, W., & Kuhn-Nentwig, L. (2019). Spider Venom: Components, Modes of Action, and Novel Strategies in Transcriptomic and Proteomic Analyses. *Toxins, 11*(10). doi: ARTN 61110.3390/toxins11100611

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research, 47*(W1), W636-W641. doi: 10.1093/nar/gkz268

Miyashita, M., Mitani, N., Kitanaka, A., Yakio, M., Chen, M., Nishimoto, S., . . . Miyagawa, H. (2021). Identification of an antiviral component from the venom of the scorpion Liocheles australasiae using transcriptomic and mass spectrometric analyses. *Toxicon, 191*, 25-37. doi: 10.1016/j.toxicon.2020.12.007

Sade, Y. B., Boia-Ferreira, M., Gremski, L. H., da Silveira, R. B., Gremski, W., Senff-Ribeiro, A., . . . Veiga, S. S. (2012). Molecular cloning, heterologous expression and functional characterization of a novel translationally-controlled tumor protein (TCTP) family member from Loxosceles intermedia (brown spider) venom. *International Journal of Biochemistry & Cell Biology, 44*(1), 170-177. doi: 10.1016/j.biocel.2011.10.013

Santos, D. M., Verly, R. M., Pilo-Veloso, D., de Maria, M., de Carvalho, M. A. R., Cisalpino, P. S., . . . de Lima, M. E. (2010). LyeTx I, a potent antimicrobial peptide from the venom of the spider Lycosa erythrognatha. *Amino Acids, 39*(1), 135-144. doi: 10.1007/s00726-009-0385-x

Undheim, E. A. B., Sunagar, K., Herzig, V., Kely, L., Low, D. H. W., Jackson, T. N. W., . . . Fry, B. G. (2013). A Proteomics and Transcriptomics Investigation of the Venom from the Barychelid Spider Trittame loki (Brush- Foot Trapdoor). *Toxins, 5*(12), 2488-2503. doi: 10.3390/toxins5122488

Wan, H., Lee, K. S., Kim, B. Y., Zou, F. M., Yoon, H. J., Je, Y. H., . . . Jin, B. R. (2013). A Spider-Derived Kunitz-Type Serine Protease Inhibitor That Acts as a Plasmin Inhibitor and an Elastase Inhibitor. *Plos One, 8*(1). doi: ARTN e5334310.1371/journal.pone.0053343

Yamazaki, Y., Hyodo, F., & Morita, T. (2003). Wide distribution of cysteine-rich secretory proteins in snake venoms: Isolation and cloning of novel snake venom cysteine-rich secretory proteins. *Archives of Biochemistry and Biophysics, 412*(1), 133-141. doi: 10.1016/S0003-9861(03)00028-6

Yans, L. Z., & Adams, M. E. (1998). Lycotoxins, antimicrobial peptides from venom of the wolf spider Lycosa carolinensis. *Journal of Biological Chemistry, 273*(4), 2059-2066. doi: DOI 10.1074/jbc.273.4.2059

Yuan, C. H., He, Q. Y., Peng, K., Diao, J. B., Jiang, L. P., Tang, X., & Liang, S. P. (2008). Discovery of a Distinct Superfamily of Kunitz-Type Toxin (KTT) from Tarantulas. *Plos One, 3*(10). doi: ARTN e3414
10.1371/journal.pone.0003414

CHAPTER 4

Biological evaluation of linear peptides identified by transcriptome analysis.

4.1 Introduction

Two major components are responsible for most of the toxic effects of the spider venom; disulfide-rich (SS-rich) peptides and linear peptides. SS-rich peptides are the dominant components in most spider venoms, which contribute to the venom's insecticidal activity as a neurotoxin by acting on ion channels(Kuhn-Nentwig, 2021; Kuhn-Nentwig, Stocklin, & Nentwig, 2011; Rohou, Nield, & Ushkaryov, 2007). On the other hand, linear peptides show various activities, such as antimicrobial activity, by interrupting the cell membrane integrity. They are characterized by formation of an amphipathic cationic α-helical structure. One of the major groups of linear peptides from spider venom is cupiennins, which were identified from the venom of *Cupiennius salei* (Ctenidae)(Kuhn-Nentwig et al., 2002). They can be divided into two families based on the presence and absence of Pro residues(Kuhn-Nentwig, Schaller, & Nentwig, 2004; Kuhn-Nentwig, Sheynis, Kolusheva, Nentwig, & Jelinek, 2013). Linear peptides in spider venoms, including cupiennins, are also known to show insecticidal activity and/or enhancing effects on neurotoxin actions in the venom(Kuhn-Nentwig, et al., 2002; Kuhn-Nentwig, et al., 2004). Due to these characteristics, there has been growing interest in the bioactive linear peptides derived from spider venoms.

Antimicrobial peptides (AMPs) are known to widely exist in nature and play an important role in the innate immune system of various organisms. They are characterized by a wide range of inhibitory effects on the growth of bacteria, fungi, parasites, and viruses(Kuhn-Nentwig, et al., 2011). As reported by WHO, reduced effectiveness of antibiotics due to the emergence of resistance to them is becoming a global risk. To counter this threat, effective substitutes for conventional antibiotics are urgently needed. Because AMPs act on biological membranes, they are less likely to emerge resistance to these peptides in these microorganisms. Therefore, AMPs can be an alternative to the conventional antibiotics, and in particular, AMPs derived from spider venom are attracting attention.

In the previous chapter, 17 linear sequences were identified as a cytolytic peptide by transcriptome analysis of the venom gland of *Lycosa poonaensis*. In this chapter, the biological activities of these linear peptides were evaluated to estimate their functions in the venom.

4.2 Material and methods

4.2.1 Peptide synthesis.

Peptides were synthesized by the Fmoc-based solid-phase method as previously described in chapter 2. Peptides were synthesized using automated microwave peptide synthesizer (Liberty Blue, CEM), in which DIC and Oxyma Pure were used for coupling. For the synthesis of C-terminal amidated peptides, Rink Amide ProTide Resin (LL) resin (CEM) were used. The synthesiezed peptides were purified by reversed-phase HPLC (RP-HPLC) on a C18 column and the identity were confirmed by LC/MS analyses.

4.2.2 Mass spectrometric analysis

LC/MS measurements were carried out on an LCMS-2020 mass spectrometer (Shimadzu, Kyoto, Japan). A RP-HPLC column (Poroshell 120-EC-C18 2.1× 75 mm, Agilent, USA) was used for separation. The column was eluted with 0.1% formic acid in H_2O (solvent A) and 0.1% formic acid in CH_3CN (solvent B) at a flow rate of 0.3 mL/min using a linear gradient of 5%-60% of solvent B over 15 min.

4.2.3 Bioassay

Antibacterial, insecticidal, and hemolytic activities were evaluated as previously described in chapter 2.

4.2.4 Circular dichroism measurements

Circular dichroism measurements was carried out as previously described in chapter 2.

4.3 Results and Discussion

4.3.1 Selection of peptides

The peptides synthesized for biological characterization were selected as follows. The calculated molecular masses of the linear peptides identified in chapter 3 were searched from the data obtained by LC/MS analysis of the crude venom in chapter 1. As a result, existence of five peptides in the venom was confirmed, which were synthesized for biological characterization. In addition, two other peptides were selected from a family different from those of the peptides selected above (Table 4-1).

Table 4-1. Sequence of the synthesized peptides. Peptides that were not detected by LC/MS analysis of the crude venom are underlined.

Name	Sequence	Molecular mass	Net charge
M-Lycotoxin-Lp1a	MVWLLPLKFLASHVAMEQLSKLGSKIATKL-NH_2	3350.9	+4
M-Lycotoxin-Lp1b	VIWLPALKFLASHIAMEQLSKL-NH_2	<u>2506.4</u>	+2
M-Lycotoxin-Lp3a	IWLTALKFIGKNLGKHFAKQQLSKL-NH_2	2880.7	+6
M-Lycotoxin-Lp4	KIKWFKAMKSIAKYVAKKQLKKHLGGEN	<u>3271.9</u>	+8
M-Lycotoxin-Lp5a	AGLGKIGAFIKKAYAIYKAKAA-NH_2	2251.3	+6
M-Lycotoxin-Lp5b	AGLGKIGALIKKAYAIYKAKAA-NH_2	2217.3	+6
M-Lycotoxin-Lp6	SKWKAFLAKMKEIASEAL-NH_2	2049.1	+3

4.3.2 Circular dichroism

All synthesized peptides were analyzed by CD spectroscopy to investigate their secondary structure in solution. The result demonstrates that all these peptides are assumed to adopt an α-helical structure in 50% trifluoroethanol, a helix-promoting solvent (Table 4-2, Figure 4-1). In addition, based on the helical wheel projection, these peptides adopt an amphipathic nature.

Table 4-2 Estimation of secondary structures of the peptides synthesized in this study

Peptides	Solvents [a)]	Percentage of secondary structure [b)]			
		α-helix	β-sheet	Turns	Unordered
M-lycotoxin-Lp1a	I	15	26	25	35
	II	61	10	11	19
M-lycotoxin-Lp1b	I	10	31	26	33
	II	57	9	12	22
M-lycotoxin-Lp3a	I	11	29	26	34
	II	59	16	8	16
M-lycotoxin-Lp4	I	8	33	26	33
	II	51	11	13	23
M-lycotoxin-Lp5a	I	15	23	28	33
	II	55	17	10	18
M-lycotoxin-Lp5b	I	7	33	25	34
	II	51	14	11	23
M-lycotoxin-Lp6	I	20	22	25	33
	II	55	14	8	22

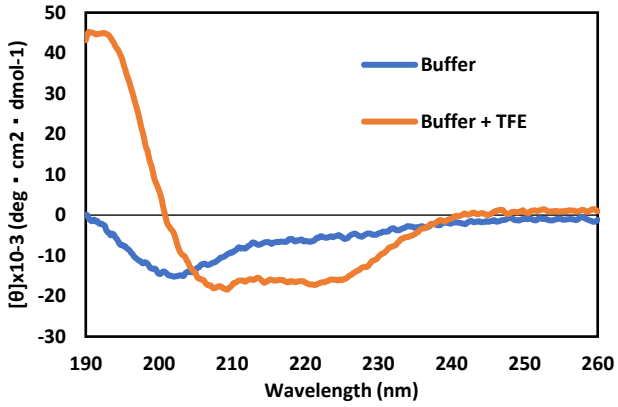
[a)] Solvent I indicates 0.2 M phosphate buffer (pH7.0), and solvent II indicates 50% TFE in 0.2 M phosphate buffer (pH7.0).
[b)] Calculated using the DichroWeb server with the CDSSTR method and reference data set 4.

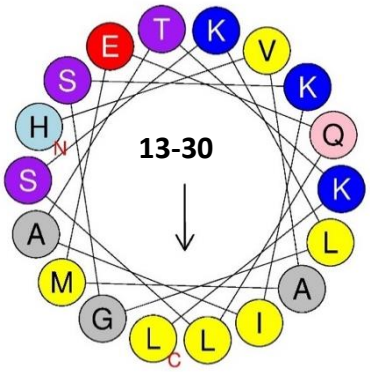
4.3.3 Biological activities

First, antibacterial activity was evaluated using Gram-negative (*E. coli*) and Gram-positive (*S. aureus* and *B. subtilis*) bacteria. As shown in Table 4-3, all peptides showed significant antibacterial activity against three bacteria, except for M-lycotoxin-Lp6, which showed no activity against *S. aureus*. When comparing the activity against *E. coli* among these peptides, M-lycotoxin-Lp4 showed the highest activity, whereas M-lycotoxin-Lp1, M-lycotoxin-Lp1b, and M-lycotoxin-Lp6 showed relatively low activity. This difference in activity is thought to be due to differences in the net charge of these peptides(Hancock & Patrzykat, 2002; Irazazabal et al., 2016). M-Lycotoxin-Lp4 has a net charge of +8, whereas M-lycotoxin-Lp1a M-lycotoxin-Lp1b, and M-lycotoxin-Lp6 has only +4, +2, and +3, respectively. Then, comparing the activity against *S. aureus*, M-lycotoxin-Lp3a has the highest activity, while M-lycotoxin-Lp5a, M-lycotoxin-Lp5b, and M-lycotoxin-Lp6 have relatively low activity. Since M-lycotoxin-Lp3a, M-lycotoxin-Lp5a, and M-lycotoxin-Lp5b have the same net charge (+6), this difference is due to other reasons. When comparing the position of the basic Lys residues on the hydrophilic face of their amphiphilic α-helical structure, Lys residues are evenly distributed in the case of M-lycotoxin-Lp3a, whereas Lys residues are located near the boundary of the hydrophilic surface in the cases of M-lycotoxin-Lp5a and M-lycotoxin-Lp5b (Figure 4-1). This difference may affect the activity against *S. aureus*. In the case of M-lycotoxin-Lp6, the net charge is also lower (+3) in addition to this difference, which is likely to result in much lower activity. When comparing the activity against *B. subtilis*, M-lycotoxin-Lp1, M-lycotoxin-Lp1b, and M-lycotoxin-Lp6 are less active than the other peptides. This may reflect differences in net charge as observed for activity against *E. coli*. However, M-lycotoxin-Lp4 did not show high activity as expected from its high net charge (+8). Other factors, such as the presence of acidic amino acid residues on the hydrophilic face, may affect its activity(Irazazabal, et al., 2016; Pathak et al., 1995).

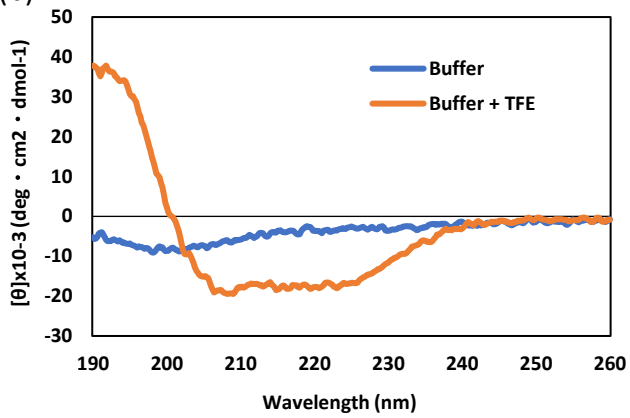
Next, the toxicity of these peptides to insects was examined (Table 4-3). All peptides showed toxicity against crickets. Among them, M-lycotoxin-Lp4 showed the highest activity with an ED_{50} value of 20 μg/g, while M-lycotoxin-Lp6 showed the lowest activity (ED_{50} = 94 μg/g). This result may reflect differences in the net charge between these peptides. However, other factors seem to be involved, because Lp1a and Lp1b, which have a lower net charge, also exhibit relatively high toxicity. These peptides contain a Pro residue, which may affect their activity by causing a bent structure.

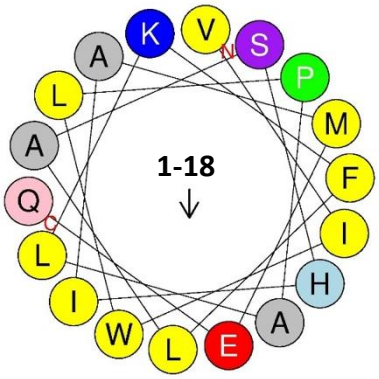
(a)


Buffer
Buffer + TFE
[θ]x10-3 (deg • cm2 • dmol-1)
Wavelength (nm)

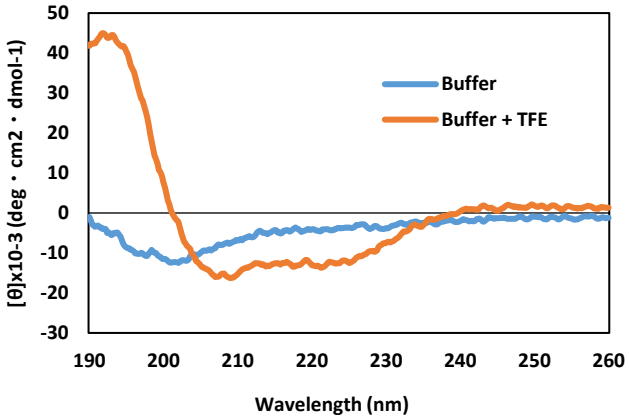

13-30

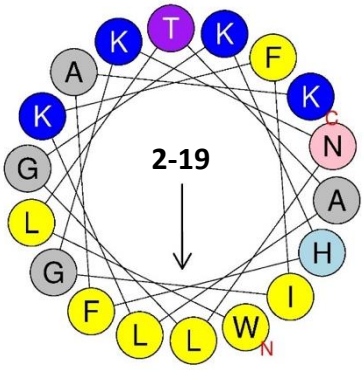
(b)


Buffer
Buffer + TFE
[θ]x10-3 (deg • cm2 • dmol-1)
Wavelength (nm)

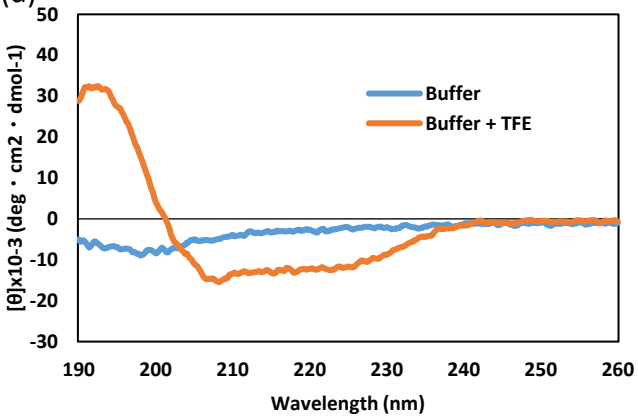

1-18

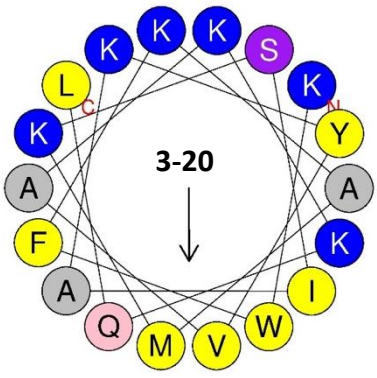
(c)

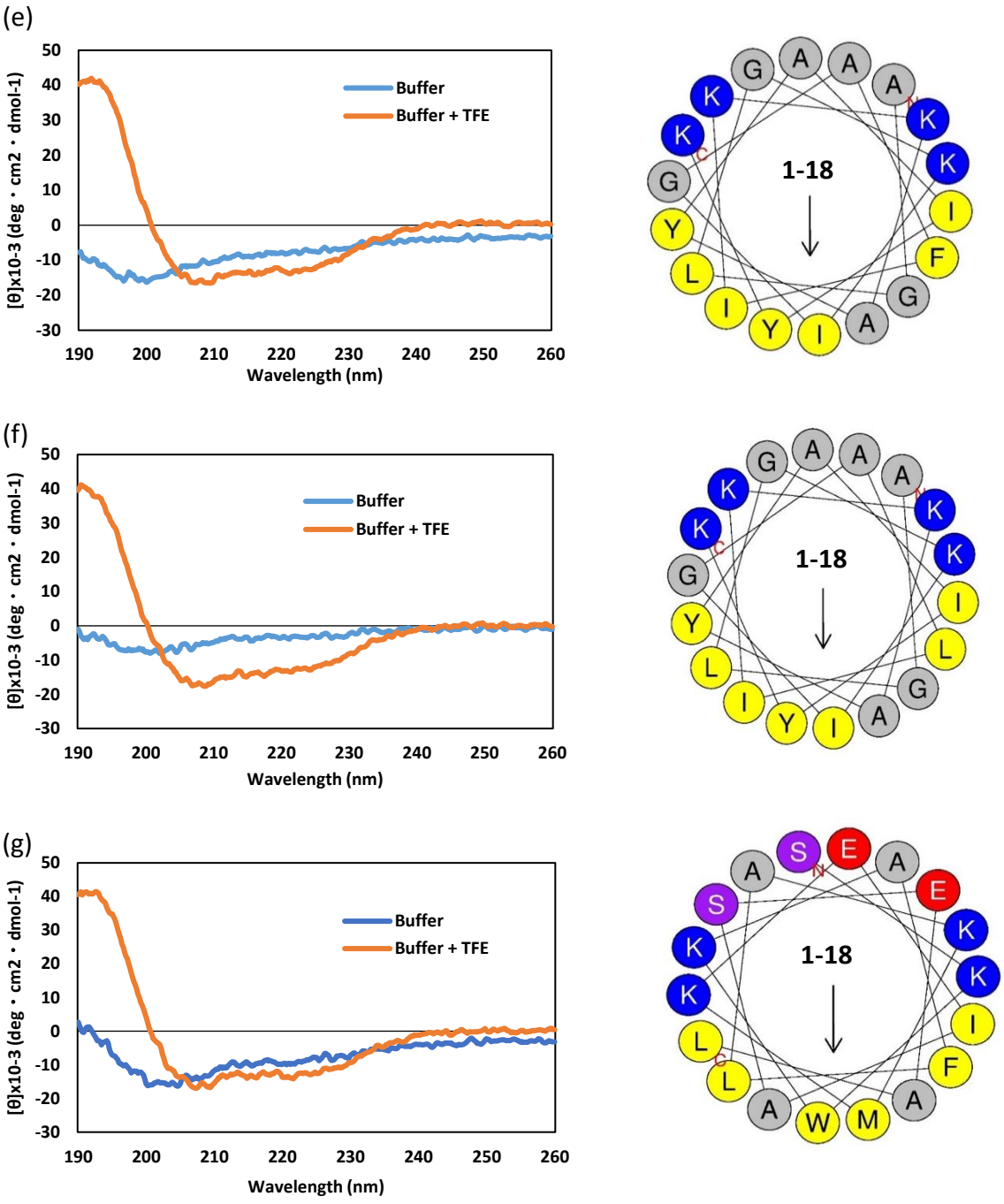

Buffer
Buffer + TFE
[θ]x10-3 (deg • cm2 • dmol-1)
Wavelength (nm)


2-19

(d)


Buffer
Buffer + TFE
[θ]x10-3 (deg • cm2 • dmol-1)
Wavelength (nm)


3-20



Figures 4-1 Secondary structure analysis of M-lycotoxin-Lp1a (a), M-lycotoxin-Lp1b (b), M-lycotoxin-Lp3a (c), M-lycotoxin-Lp4 (d), M-lycotoxin-Lp5a (e), M-lycotoxin-Lp5b (f), and M-lycotoxin-Lp6 (g). CD spectra were obtained in 0.2 M phosphate buffer at pH 7.0 (dotted line) and in 50% TFE in 0.2 M phosphate buffer at pH 7.0 (solid line). Helical wheel projections were obtained using HeliQuest. Arrows indicate hydrophobic moments.

Table 4-3. Biological activity of the seven representative peptides.

Peptide	Antibacterial activity MIC (μM)			Insect toxicity ED_{50} (μg/g weight)	Hemolytic activity (μM)
	E. coli	*S. aureus*	*B. subtilis*		
M-Lycotoxin-Lp1a	12.5	25	25	34	>100
M-Lycotoxin-Lp1b	100	6.2	12.5	30	>100
M-Lycotoxin-Lp3a	1.6	0.8	1.6	70	>100
M-Lycotoxin-Lp4	0.7	3.1	6.2	10	>100
M-Lycotoxin-Lp5a	2.5	100	2.5	50	>100
M-Lycotoxin-Lp5b	2.5	100	2.5	44	>100
M-Lycotoxin-Lp6	12.5	>100	12.5	94	>100

Furthermore, hemolytic activity of these peptides was examined (Table 4-3). As a result, none of them showed activity even at 100 μM. It has been reported that hydrophobicity of peptides, rather than net charge, is important for expression of hemolytic activity(Dathe et al., 1997). This is because mammalian cell membranes are composed of mostly zwitterionic phospholipids and cholesterol, making the membrane neutral and hydrophobic. The peptides used in this study may be relatively less hydrophobic so that they are not able to interact with the erythrocyte membranes.

M-Lycotoxin-Lp5a and M-lycotoxin-Lp5b have similar sequences, differing only the 9th residue (Phe or Leu). These two peptides showed the similar activity profile, indicating that the difference of this amino acid residues does not affect the activity.

In chapter 2, it was shown that lyp2370 has only weak antibacterial activity, although it adopt an amphiphilic α-helical structure typical of antimicrobial peptides (Table 4-4). As described in chapter 3, lyp2370 shares a common sequence with M-lycotoxin-Lp1a, but lacks its N-terminal region (MVWLLPLK).

Since M-lycotoxin-Lp1a showed significant antibacterial activity and insect toxicity, the low activity of lyp2370 is due to the lack of the N-terminal region from M-lycotoxin-Lp1a. The lacking N-terminal sequence in lp2370 contains many hydrophobic residues. In addition, there is a Pro residue that can interfere with the formation of an α-helical structure. Thus, the hydrophobic N-terminal region in M-lycotoxin-Lp1a is likely to contribute to activity in a different manner from the main α-helical structure.

Table 4-4. Comparison of biological activities between lyp2370 and M-lycotoxin-Lp1a.

Peptide (sequence)	Antibacterial activity MIC (μM)			Insect toxicity ED_{50} (μg/insect)
	E. coli	*S. aureus*	*B. subtilis*	
lyp2370 (**FLASHVAMEQLSKLGSKIATKL**-NH_2)	>200	>200	200	>1000
M-Lycotoxin-Lp1a (**MVWLLPLKFLASHVAMEQLSKLGSKIATKL**-NH_2)	12.5	25	25	34

4.4 Summary

In summary, from the 17 linear peptides identified, seven representative peptides of each family were chemically synthesized, and their biological activities were evaluated. All peptides showed significant antibacterial activity against Gram-negative (*E. coli*) and Gram-positive (*S. aureus and B. subtilis*) bacteria, although their selectivity for bacterial species differed. Some peptides also showed toxicity against insects (crickets), but none of them showed hemolytic activity. The secondary structure analysis by circular dichroism spectroscopy shows that all these peptides form an α-helical structure. Furthermore, these peptides are presumed to have an amphiphilic structure with distinct hydrophilic and hydrophobic surfaces in the helix structure. The differences in their activities observed in this study are possibly due to the net charge, the arrangement of basic and acidic residues, and the hydrophobicity of the peptides.

References

Dathe, M., Wieprecht, T., Nikolenko, H., Handel, L., Maloy, W. L., MacDonald, D. L., . . . Bienert, M. (1997). Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides. *FEBS letters, 403*(2), 208-212.

Hancock, R., & Patrzykat, A. (2002). Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. *Current drug targets-Infectious disorders, 2*(1), 79-83.

Irazazabal, L. N., Porto, W. F., Ribeiro, S. M., Casale, S., Humblot, V., Ladram, A., & Franco, O. L. (2016). Selective amino acid substitution reduces cytotoxicity of the antimicrobial peptide mastoparan. *Biochimica Et Biophysica Acta-Biomembranes, 1858*(11), 2699-2708. doi: 10.1016/j.bbamem.2016.07.001

Kuhn-Nentwig, L. (2021). Complex precursor structures of cytolytic cupiennins identified in spider venom gland transcriptomes. *Scientific Reports, 11*(1). doi: ARTN 400910.1038/s41598-021-83624-z

Kuhn-Nentwig, L., Muller, J., Schaller, J., Walz, A., Dathe, M., & Nentwig, W. (2002). Cupiennin 1, a new family of highly basic antimicrobial peptides in the venom of the spider Cupiennius salei (Ctenidae). *Journal of Biological Chemistry, 277*(13), 11208-11216. doi: 10.1074/jbc.M111099200

Kuhn-Nentwig, L., Schaller, J., & Nentwig, W. (2004). Biochemistry, toxicology and ecology of the venom of the spider Cupiennius salei (Ctenidae). *Toxicon, 43*(5), 543-553. doi: 10.1016/j.toxicon.2004.02.009

Kuhn-Nentwig, L., Sheynis, T., Kolusheva, S., Nentwig, W., & Jelinek, R. (2013). N-terminal aromatic residues closely impact the cytolytic activity of cupiennin 1 a, a major spider venom peptide. *Toxicon, 75*, 177-186. doi: 10.1016/j.toxicon.2013.03.003

Kuhn-Nentwig, L., Stocklin, R., & Nentwig, W. (2011). Venom Composition and Strategies in Spiders: Is Everything Possible? *Advances in Insect Physiology, Vol 40: Spider Physiology and Behaviour - Physiology, 40*, 1-86. doi: 10.1016/B978-0-12-387668-3.00001-5

Pathak, N., Salas‐Auvert, R., Ruche, G., Janna, M. h., McCarthy, D., & Harrison, R. G. (1995). Comparison of the effects of hydrophobicity, amphiphilicity, and α‐helicity on the activities of antimicrobial peptides. *Proteins: Structure, Function, and Bioinformatics, 22*(2), 182-186.

Rohou, A., Nield, J., & Ushkaryov, Y. A. (2007). Insecticidal toxins from black widow spider venom. *Toxicon, 49*(4), 531-549. doi: 10.1016/j.toxicon.2006.11.021

CHAPTER 5

Summary of the thesis

Spiders developed about 400 million years ago, and they are one of the most diverse and successful terrestrial organisms with more than 50,000 species. Spider venom is a complex mixture of bioactive components, in which peptides play an important role which displays multiple functions, such as neurotoxic, analgesic, cytotoxic, necrotic activities. Among them, linear peptides have important functions in spider venom. They are mainly responsible for disrupting the integrity of the cellular membrane of bacteria and preventing the infection of pathogens during injection of venom into prey or ingestion of prey.

In this study, the venom of the spider *L. poonaensis* was analyzed using mass spectrometry to find non-disulfide-bridged peptides as an AMP candidate. The result revealed that 120 out of 401 components were non-disulfide-bridged peptides. The amino acid sequences of two peptides (lyp2370 and lyp1987) were determined by MS/MS analysis, confirmed by comparison with synthesized samples. I also found that two other peptides (lyp2130 and lyp1875) have C-terminally truncated lyp2370 and lyp1987, respectively. The antibacterial activity test revealed that lyp2370 has only weak activity, whereas lyp1987, which has a structure identical to M-lycotoxin-Ls3b, showed significant activity. The lower activity of lyp2370 was due to the presence of a Glu residue on the hydrophilic face of its amphiphilic α-helical structure.

Then, the transcriptomics analysis of the *Lycosa poonaensis* venom gland was carried out in this study. This resulted in the identification of 87 transcripts, representing peptides and proteins similar to those of other spider venoms. The number of sequences identified as neurotoxins was the highest, followed by peptidases and cytolytic toxins that consist of linear peptides without disulfide bonds. As expected from the results of mass spectrometric analysis, it was confirmed that this venom contains a large number of linear peptides, which likely show antimicrobial activity.

From the 17 linear peptides identified by the transcriptomics analysis, seven representative peptides of each family were chemically synthesized, and their biological activities were evaluated. Although their selectivity for bacterial species differed, all peptides showed significant antibacterial activity against Gram-negative (*E. coli*) and Gram-positive (*S. aureus* and *B. subtilis*) bacterial species. Some peptides also showed toxicity against insects (crickets), but none of them showed hemolytic activity.

The secondary structure analysis by circular dichroism spectroscopy shows that all these peptides form an α-helical structure. Furthermore, these peptides are presumed to have an amphiphilic structure with distinct hydrophilic and hydrophobic surfaces in the helix structure, suggesting that they exert their activity by disrupting the cell membrane structure of bacteria and insects.

List of publication

1. **Megaly AMA**, Yoshimoto Y, Tsunoda Y, Miyashita M, Abdel-Wahab M, Nakagawa Y, Miyagawa H., Characterization of 2 linear peptides without disulfide bridges from the venom of the spider *Lycosa poonaensis* (Lycosidae). Biosci Biotechnol Biochem. 2021 May 25;85(6):1348-1356. doi: 10.1093/bbb/zbab046. PMID: 33729438.

2. **Megaly AMA**, Miyashita M, Abdel-Wahab M, Nakagawa Y, Miyagawa H., De novo transcriptome analysis of the venom gland of the spider *Lycosa poonaensis* (Lycosidae), in preparation.

3. **Megaly AMA**, Miyashita M, Abdel-Wahab M, Nakagawa Y, Miyagawa H., Biological evaluation of linear peptides identified by transcriptome analysis of the venom gland of the spider *Lycosa poonaensis* (Lycosidae), in preparation.

ACKNOWLEDGMENTS

In the Name of Allah, the Most Beneficent, the Most Merciful.

First, I am thankful to ALLAH ALMIGHTY, who blessed me with sound health and abilities and gave me the courage to perform and complete my work successfully. Without the help of ALLAH, I could not have had the ability to do my work completely.

I want to express my deepest appreciation to Prof. Hisashi Miyagawa for accepting me in his laboratory, Bioregulation Chemistry, Division of Applied Life Science, Graduate School of Agriculture, Kyoto University. Also, I am very thankful for his guidance, support, and recommendation through my study and living in Kyoto.

I am very grateful to Associate Prof. Yoshiaki Nakagawa for his astonishing kind and support. He is doing his best and spent time, effort to support and help in all possible respects.

I would like to deeply thank Assistant Prof. Masahiro Miyashita for his guidance, support, discussion, and recommendation through my study or even in different aspects of life.

Because of our daily contact and discussion, his way of thinking and judgment on the scientific field and life aspects had a huge effect on me. Honestly, I am very glad to meet and work with him.

I cannot forget Dr. Mohamed Abdelwahab, my mentor and one who established the relationship with the Bioregulation chemistry laboratory and strongly recommended me form the first place.

Also, I would like to express my thanks to all the laboratory members for their kindness and supports throughout this study.

My deep love and special thanks to my wife and daughters (Fayroze and Farida) and my deepest appreciation to my elder brothers and sisters.

Finally, I would like to dedicate this work to the soul of my parents (May ALLAH accept them in heaven).