Identification of < 10 KD peptides in water extraction of Venenum Bufonis from Bufo gargarizans using Nano-LC-MS/MS and De novo sequencing

Identification of ＜10KD peptides in the water extraction of *Venenum Bufonis* from *Bufo gargarizans* using Nano LC-MS/MS and *De novo* sequencing

Yiguo Huo[1,2], Ruoxian Xv[1,], Hongyue Ma[1,*], Jing Zhou[1], Xinping Xi[2,*], Qinan Wu[1], Jinao Duan[1], Mei Zhou[2], Tianbao Chen[2]

[1]Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, and Jiangsu Key Laboratory for High Technology Research of TCM Formulae, College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; [2]Natural Drug Discovery Group, School of Pharmacy, Queen's University, Belfast BT9 7BL, Northern Ireland, UK.

* Correspondence: hongyuema@126.com (H.M.); x.xi@qub.ac.uk (X.X.)

Abstract

Skins of anurans (frogs and toads) are rich sources of bioactive peptides. However, the peptides secreted by the skin glands of *Bufo gargarizans*, the most common toad in China, remained unexplored to date. Here, a strategy combines LC-MS/MS, RNA sequencing and bioinformation analysis was applied to unravel the peptides in the *Bufo gargarizans* secretions. Data-dependent LC-MS/MS acquisitions of intact peptides followed by automated chromatographic alignment, *De novo* analysis, database and homology searches with manual validations showed that the venom is composed by 939 features, with masses ranging from 0.7−4 kDa. These peptides derived from 85 proteins were identified using the PEAKS software with acquired MS and MS/MS spectra of *Venenum Bufonis* against the house-built protein database using *De novo* RNA sequencing, while only 23 peptides from 8 proteins were found

when searching known amphibian database. Moreover, it was found that many peptides with high abundance in *Venenum Bufonis* derived from proteolytic processing of a larger precursor protein, named as CL4590. Molecular cloning was applied to validate a short domain of CL4590 to evident the accuracy of these obtained sequences. Although the bioactivities of peptides identified by MS/MS are unknown, the next function annotation showed that they may involve in the cell killing, immune and metabolic process, antioxidant activity and antimicrobial actions. Therefore, the peptidomics analysis on *Bufo gargarizans* discover abundant novel toad peptides which broaden our horizons on the secretion multiplicity and supplied an assortment of pharmacological candidates.

key words *Bufo gargarizans*; Peptidomics; Mass Spectrometry; *De novo* sequencing; Molecular Cloning

1. Introduction

Amphibians have evolved various defense strategies for survival. Their bioactive peptides, especially the antimicrobial and anticancer peptides from the secretions, have attracted much attention during the past decades. To date, more than a thousand bioactive peptides from amphibian have been deposited into the Antimicrobial Peptides Database (http://aps.unmc.edu/AP/main.php) and Anticancer Peptides Database (http://crdd.osdd.net/raghava/cancerppd/index.php). In addition, many antimicrobial peptides have been found to have antitumor effects [1]. Nowadays, secretion of several species (snake, sorption, toad, frog) and their active components (protein, peptides, enzymes) have revealed various therapeutic effect. PGLa, an antimicrobial peptide from the magainin family, has been found in the skin of *Xenopus laevis*. It demonstrates a wide spectrum activity against Gram-positive and Gram-negative bacteria [2]. Melittin, a 28-mer α-helical peptide from bee venom, inhibited several drug resistant bacteria in vitro and degraded the biofilm layer [3]. A peptide from the venom of *Buthus Martensii Karsch* inhibited the propagation of fibrosarcoma cells [4]. Magainin 2 could lyse hematopoietic tumor rapidly, without damaging normal cells [5]. Bengalin, a protein separated from Indian black scorpion

(*Heterometrus bengalensis*) venom, showed antiproliferative activity on leukemic cell lines [6]. A neuropeptide, isolated from the venom of Stony Creek Frog *Litoria lesueuri*, showed anticancer activity on 60 different human cancer cells [7]. Antibacterial and anticancer studies on extraction from animal secretions are explored all over the world. However, research on proteins and peptides from the secretion of Chinese toad (*Bufo gargarizans*) is relatively insufficient. This important natural property of antimicrobial or antitumor peptides to be potential drugs represents an excellent value for therapeutic research.

Bufo gargarizans, one of the biggest species of *Bufonid* in China, whose secretion contains various small molecules with significant anticancer activity, has drawn much attention so far [8]. The parotid secretion of *Bufo gargarizans*, *Venenum Bufonis*, has been used in clinic as a traditional Chinese medicine for nearly 2000 years [9], and it is well-known because of its anticancer activity as well as toxicity. It has been widely acknowledged that its efficient and poisonous components were some small molecules such as bufalin and cinobufagin [10,11]. However, the anticancer concentration of small molecules *in vivo* resulted in serious cardiotoxicity [12]. While two peptides from *Bufo gargarizan*s and their derivations showed remarkable antimicrobial and antitumor action. One of which is Buforin, which was discovered in the stomach mucosa of the Asian toad *Bufo gargarizans* [13] and the other is a novel Cathelicidin termed BG-CATH, which was identified by analysis of the toad skin transcriptome [14].

Nowadays, identifying small peptides is a great challenge because mass spectrometry tools are not permanently appropriate for the detection of molecules within small mass range. Usually, fragmentation of small peptides produces low peaks with poor responsibility which lead to vague identified sequences [15]. The enormous complication and considerable number of peptides result in the extreme challenge to investigation on animal secretions. Nevertheless, studying the comprehensive multiplicity of secretion and toxicities in venom is considerable because of the massive potential substances in animal venoms [16]. As earlier studies in articles, there are many useful methods for identification of peptide at transcriptomic, peptidomics or genomic level [17]. Despite the increase in the number of reports on venom peptidomics, the comprehensive knowledge of the intricacy and dynamic range of

proteins present within animal venom remains a great challenge [18]. Furthermore, there are lots of limitations of present technique, the combination strategy is necessary for identify peptides with high-efficiency [19,20]. Currently, the methods are improved by the combination of peptidomics and transcriptomic studies [21,22], which could successfully unmask the composition of animal secretion. Here, we explored the combination technique to reveal the peptidomic constitution of one of the most important animal-venom-derived Chinese medicine, *Bufo gargarizans*, which could improve the understanding of the molecular biological mechanisms of *Bufo gargarizans* in the treatment of human disease.

2. Materials and Methods

2.1 Reagents and materials

DL-Dithiothreitol (DTT), Iodoracetamide (IAA), penicillin-streptomycin solution, fetal bovine serum (FBS) and MTT powder were purchased from Sigma (St. Louis, MO, USA). RPMI-1640 culture medium and DMSO were supplied by Invitrogen (Paisley, UK). Water was purified using a Mil-lipore Milli-Q system (Bedford, MA, USA). Acetonitrile and Formic acid were bought from TEDIA (Fairfield, USA) The SMMC-7721 cell lines were obtained from Shanghai Cell Institute, China Academy of Sciences.

2.2 Animals

Specimens of Chinese toad, *Bufo gargarizans* (n=3) were obtained in Jiangsu Province in China. The parotid secretion was obtained via mild squeezing and massaging of the glands. The viscous white parotid secretion was washed using deionized water, snap-frozen in liquid nitrogen, lyophilized and stored at -20 °C prior to analysis.

2.3 Venom extraction

200 mg parotid secretion extracted from *Bufo gargarizans* were dissolved in 1.5 ml PCR water and then ultrasonicated by a ultrasonication for 6 min and extracted by ultrasonic for another 30 min by ice-bathing. After that the tubes were incubated in an incubator at 4 °C for 5 h. Next, the supernatant (about 1.5 ml) were divided in two parts and transferred drop by drop onto the center of 10 KD molecular weight cut-off

separately. Afterwards, these two filter cartridges were centrifuged at a speed of 8000 r/min for 20 min. Subsequently, combined the products by completely pipetting. 200μl filtered products was allocated into two well-prepared PCR tubes. And 12.5 μl DTT was added in one PCR tube at room temperature. Then, 7.5 μl IAA was also added in the tube in a dark environment for 45 min. Finally, all the samples were desalted by oasis HLB columns (Waters, USA) and dried by a concentrator (Eppendorf, Hamburg, Germany) at 45 °C.

2.4 Anticancer activity of total peptides on human hepatocarcinma cell lines

The human hepatocarcinma cell lines (SMMC-7721) were maintained in RPMI-1640 culture medium. RPMI-1640 were supplemented with 1% penicillin-streptomycin solution and 10% fetal bovine serum. The cells were placed in an incubator (Stuart, UK) with a 90% humidified environment containing 5% CO_2 at 37 ^oC. The cell line was prepared at the density of 5×10^3 cells per well onto 96 well plates. Following this, cell lines were loaded with peptides at final concentrations ranged between 0.01mg/ml~500μg/ml and incubated over 24 h. After this, 10 μL of 5 mg/mL MTT solution were added into each well and incubated for 4 h. The supernatants were removed by a syringe, followed by addition 100 μL of 1% DMSO. The absorbance of each well was recorded by PowerWaveX340 (BioTek instruments, USA) at 490 nm.

2.5 Mass spectrometry

Each sample was resuspended in 0.1%FA and centrifuged at 14000 g for 5 min. The supernatant (4 μl) was loaded onto a 2 cm Thermo column (PepMap, 75 μm*2 cm C18, 3 μm, 100 Å) (Dionex, USA) with a Nano-LC system (Dionex UltiMate 3000) (Thermo, USA) and the peptides were eluted onto a resolving 15 cm analytical C18 column (Acclaim PepMap RSLC, 50 μm*15 cm C18, 2 μm, 100Å) (Dionex, USA). A gradient elution of solvent A (formic acid: acetonitrile: water, 0.1/4/95.9) and solvent B (formic acid: water: acetonitrile, 0.1/4/95.9) was applied as follows: 0~10 min, 1% B; 10~12 min, 1%~3% B; 12~102 min, 3% ~ 20% B; 102 ~116 min, 20%~30% B; 116~120 min, 30%~90% B; 120~128 min, 90%B; 128~129 min, 90%~1% B; 129~150 min, 1% B. The peptides were performed to a Nano-electrospray ionization followed by tandem

mass spectrometry (MS/MS) in an LTQ Orbitrap Velos Pro (Thermo, USA) coupled online to the HPLC. The mass spectrometry was operated in data-dependent mode, automatically switching between MS and MS2 acquisition. Survey full-scan MS spectra (m/z 350 –1800) were acquired in the Orbitrap with a resolution of 60,000. The 15 most intense ions were sequentially isolated and fragmented by collision induced dissociation (CID). Peptides with unassigned charge states, as well as less than +2, or more than +6, were excluded from fragmentation. Fragment spectra were recorded in the Orbitrap mass analyzer with resolution of 17,500. The dynamic exclusion was enabled with repeat count two, and an exclusion duration 20 s.

2.6 RNA sequencing

A 5 mg sample of the lyophilized bufo parotid secretion powder was dissolved in 1ml of cell lysis buffer (Dynal Biotech, UK). Polyadenylated mRNA was extracted from it by Dynabeads® mRNA DIRECT™ Kit (Ambion, Life Tech, Oslo, Norway), then subjected to cDNA library construction procedure by using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England BioLabs Inc, UK). According to the instructions, the cDNA library construction consisted of these main steps. First of all was the mRNA fragmentation by incubating at 94 °C with random primers. Then the RNA fragments were subjected to first strand cDNA and second strand cDNA synthesis. After purification with 1.8×Agencourt AMPure XP beads, and repair of the cDNA library was performed by incubating with the end repair reaction buffer, the nucleotide A was added to the 3' end of the DNA fragments to avoid ligated to each other. Therefore, the adapters with the nucleotide T at the 3' end were ligated to the DNA fragments. At last, the DNA fragments with adapters were enriched by PCR reaction and purified by AMPure XP beads. The quality of the cDNA library was verified by using Agilent 2100 Bioanalyzer (Agilent Technologies, USA) with Agilent DNA 1000 kit (Agilent Technologies, USA). The quantity of the cDNA library was validated by qPCR with KAPA SYBR® FAST qPCR kit (KAPA Biosystems，USA).

The cDNA library was then loaded into a flow cell with oligos complementary to the adapters to generate clusters through bridge amplification. Finally, the transcriptome was obtained by doing RNA-Sequencing on the Illumina HiSeq 2000 platform. The

raw data obtained from Hiseq platform were analyzed as bellows. Firstly, the index primers used to identify different samples were removed. Secondly, the data was transferred to the program FastQC 0.1.0.1 to filter the reads with low quality and less than 25 nucleotides. The filtered data was saved as fastq files. At last, the clean reads were *De novo* assembled using the software Trinity to obtain the transcriptome. RNA-seq generated 52,028,702 sequence reads that were assembled into 90,986 transcripts. The complete transcriptomic analyses of these samples are under development and will be published elsewhere.

For the identification of the mRNA corresponding to the *Venenum Bufonis* proteins investigated in this work, Trinity contigs were subjected to a BlastX search against a compiled set of spider toxins downloaded from GenBank and Uniprot KB databases, and those showing some degree of similarity were separated in a subset. The translated amino acid sequences of this subset were used as house-built database for the MS/MS analysis.

2.7 Molecular cloning

The isolation of pure mRNA from thick parotid secretion was achieved by utilizing a magnetic oligo-dT bead kit (Dynal Biotech, Merseyside, UK) which could bind polyadenylated mRNA in the cell lysis buffer supplied with the kit. Reverse transcription and synthesis of the first-strand cDNA were followed by a 3'-RACE reaction to isolate target antimicrobial peptide precursor nucleic acid sequence data with a SMART-RACE kit (Clontech, Palo Alto, CA, USA). 3'-RACE was facilitated by a sense primer (B.G. S1: 5'-CCATTCAGACGTTCCTTCTCAGTA-3') and an antisense primer (B.G.AS1: 5'-TAGCCCTGTACTTCTCCATGTATC-3'), which were then designed for the polymerase chain reaction according to the discovered RNA-Seq data of CL4590. The PCR cycling procedure included an initial denaturation step at 94 °C maintained for 60 s, then 40 thermal cycles which involved 20 s at 94 °C for denaturation, primer annealing for 10 s at 62 °C, and 240 s for extension at 72°C. The PCR products were purified by gel electrophoresis and cloned using a pGEM-T vector system (Promega Corporation, Southampton, UK). The DNA sequences of clones were obtained by use of an ABI 3100 automated capillary sequencer (Applied Biosystems, Foster City, CA, USA).

2.8 Bioinformatic analysis

Mass spectrometry raw data of peptide fractions were submitted to database searches in PEAKS Studio using tolerance of precursor and fragment mass of 10 ppm and 0.025 Da. *De novo* peptides, whose average local confidence (ALC) scores ⩾80% were selected for database searches against the house-built transcriptome-derived database and amphibian database downloaded from NCBI with the same conditions. The peptide false discovery rate (FDR) was predictable by the decoy fusion method and was selected at a maximum of 1%. Post-translational modifications and homology searches were performed by PEAKS PTM and SPIDER, respectively. Only features detected above the intensity threshold of 150 and in at least 2 out of 3 replicates were considered for further analysis.

3. Results and discussion

Studies on animal venoms are widely conducted from all over the world [23]. However, research on proteins and peptides from *Venenum Bufonis*, secretion of *Bufo gargarizans*, is relatively insufficient. And the sequences and structures of these peptides are worth to be valued. Bioactive peptides are fragment of the innate immune system of amphibian, insects and mammals and are generally small (<10 kDa) [24].

3.1 LC-MS analysis of peptides in Venenum Bufonis

In this study, to focus on small peptides, peptides were extracted by water and filtered by 10 kDa ultrafilters. The ion map after full-scan of peptides was exhibited in Figure 1. *De novo* sequencing performed by PEAKS software generate large numbers of peptides focused on the 700–4000 Da mass range. The analysis gave out sequence data and peptide identifications in *B. gargarizans* venom peptidome at this moment. It functioned as a complementary approach to analysis generate main information of intact masses of native peptides. *De novo* analysis of peptide fragments leads to substantial sequence assignments. The number of peptides were detected with molecular weight from 0.7 to 1 kDa, 1 to 2 kDa, 2 to 3 kDa, 3 to 4 kDa were 135, 822, 77 and 2, respectively (Figure 2). The 10 highest intense ions presented main clusters of masses around 1.2 kDa, with precursor ion charges ranging from 2 to 3. Most ions

eluted between retention times of 60–116 min (Table 1), which were equal to gradients around 11 – 30% of acetonitrile.

In total, the native peptide fraction of *B. gargarizans* venom LC-MS/MS analysis followed by database, PTM and SPIDER searches resulted in a total of 939 unique peptides identified from 85 proteins against the house-built protein database using *De novo* RNA sequencing. Part of them were list in Supplemental Table 1 and Table 2. However, only 23 peptides from 8 proteins were found when searching downloaded amphibian database in Uniprot (Supplemental Table 2). The proteins identified with the highest scores and the maximum number of peptides named CL4590 from *Venenum Bufonis*. To the best of our knowledge, this is a functional unknown protein. CL4590 precursor is composed of 897 amino acids and is cleaved to produce 296 peptides detected by LC-MS/MS (Supplemental Figure 1). In addition, these intact unique peptides derived from CL4590, covering 54% of the whole protein sequence (Supplemental Figure 2). Usually there are three kinds of poly-peptide skeleton on the general fracture mode, the a/b/c are behalf of N-terminal of peptide and x/y/z are representing the C-terminal of peptide. For example, the best peptide sequence matches to the protein is M(+15.99)LANEDYASLTKGIQNL, the charge of the sequence is 2, m/z 948.9697. This can be verified by the mass of each amino acid, such as the amino acid between ion y-14 and y-13 by the presence between the ion b-4 and b-3, which corresponds to an amidated ASN. Comparing to another peptide LANEDYASLTKGIQNL, N-Terminal absence of Met and none oxidized modification rendered the observation of 147 mass decrease. The mass spectra for both peptides are displayed in Figure 3. The peptides derived from CL4590 mainly focus on four parts of the protein, Figure 4 showed one part of the protein. BLAST searches and sequence alignments exhibited the protein homology to protein NP_0010881 and all the four main parts are possibly a TRPM8 channel-associated factor (Supplemental Figure 3). TRPM8 is a member of the transient receptor potential (TRP) superfamily of cation channels, which is a calcium-channel protein. The expression of TRPM8 is strongly positive-regulated in cancer cells.

267

268 **Table 1**. The 10 highest intense ions detected in the *Bufo gargarizans* venom peptidome by LC-MS/MS analysis.

Protein Accession	Peptide	Unique	-10lgP	Mass	ppm	m/z	z	RT	Intensity
Unigene2127_All	K.TKPTDDELKEL.Y	Y	69.37	1287.656	2.7	644.8369	2	68.9	6.02E+07
Unigene2127_All	Q.AEFDKAAEDVKKLK.T	Y	98.21	1590.862	0.6	796.4386	2	68.52	5.67E+07
CL4590.Contig1_All	L.SEWEENPM(+15.99)NKYVS	Y	88.84	1627.682	0.7	814.8491	2	72.25	4.31E+07
CL4590.Contig1_All	E.WEENPMNKYVS	Y	72.48	1395.613	1.1	698.8145	2	91.03	3.29E+07
CL4590.Contig1_All	L.ANEDYASLTKGIQNL.N	Y	108.77	1635.81	-0.3	818.9122	2	114.87	2.43E+07
CL127.Contig3_All	A.TPVDWKPGDRVM.V	N	67.98	1399.692	1.6	467.572	3	86.36	2.37E+07
CL4590.Contig1_All	F.FDFPGNKIT.S	Y	79.6	1037.518	0.3	519.7665	2	96.87	1.91E+07
CL4590.Contig1_All	F.FDFPGNKITSVAGVY.F	Y	75.42	1613.809	2.1	807.9135	2	118.22	1.89E+07
CL4590.Contig1_All	G.WDPFKQVF.S	Y	70.7	1065.528	0.3	533.7716	2	123.67	1.72E+07
CL4590.Contig1_All	W.GWPIDDATTEKL.S	Y	97.75	1344.656	0.3	673.3355	2	110.63	1.54E+07

Table2. Proteins (partial) identified in the secretion of *Bufo gargarizans* by LC-MS/MS through searching transcriptome-derived database.

Accession	-10lgP	Coverage (%)	Intensity	#Peptides	#Unique	#Spec	Description
CL4590.Contig1_All	610.2	54	7.26E+08	296	296	635	minus strand unknown [*Bufo gargarizans*]
Unigene2773_All	491.62	60	2.75E+08	110	110	297	LOC733366 protein [Xenopus laevis]
Unigene2127_All	423.53	93	3.13E+08	74	74	187	minus strand diazepam binding inhibitor [*Bufo gargarizans*]
Unigene13919_All	405.09	42	9.87E+07	76	76	154	catalase [Glandirana rugosa]
CL6550.Contig2_All	364.42	21	1.16E+07	36	36	57	inner-ear cytokeratin [Rana catesbeiana]
Unigene18322_All	355.25	56	1.72E+07	36	36	52	brain abundant membrane attached signal protein 1 [Xenopus (Silurana) tropicalis]
CL7798.Contig1_All	311.05	49	7.59E+07	15	15	40	diazepam binding inhibitor [Xenopus laevis]
Unigene19123_All	266.99	89	1.59E+07	20	20	25	minus strand Predicted: protein diaphanous homolog 1 [Xenopus (Silurana) tropicalis]
Unigene12309_All	229.43	23	1.53E+06	10	10	11	Predicted: calpain small subunit 1 [Xenopus (Silurana) tropicalis]
CL2275.Contig5_All	229.09	5	2.48E+06	8	8	9	Predicted: filamin-A isoform 3 [Ornithorhynchus anatinus]

3.2 Validation of CL4590 by molecular cloning

The peptide with the largest number of unique peptides derived from CL4590 was identified, which derived from the parotid secretion of *Bufo gargarizans* and it consists of 2970 nucleotides (Supplemental Figure 2). The translated open reading frame of CL4590 contains 897 amino acids. We focused on this protein due to its substantial quantity in the secretions. Therefore, a further identification of a short domain in CL4590 at the genetic level was performed using specific primers. The band around 500 bp of PCR product is consistent with the size of target domain (Figure 5). The validated fragment, including 502 nucleotide acids, shows only 5 nucleotide acids are inconsistent with CL4590 and result in two amino acids different from the corresponding sites of CL4590 on the aspect of amino acid sequences. The CL4590 sequence established by *De novo* RNA-Sequencing had more than 99% overlap with the validated fragment using molecular cloning (Figure 6A, 6B). This illustrated that our peptide sequences identified in *Venenum Bufonis* were accurate. The full-length nucleotide sequence of CL4590 and the validated partial sequence were deposited into Genbank database under the accession numbers Banklt 2053205 Seq1 MG189609 and Banklt 2053207 Seq1 MG189610, respectively.

The transcriptome method based on RNA-seq reveals a powerful implement to improve the number of the peptides identification. RNA-Seq reveals the presence of RNA in a biological sample at a given moment in time. RNA-Seq eliminated several challenges raised by microarray technologies due to the limitation of dynamic range of detection. However, during the procedure of reverse transcription, the emerging cDNA was synthesised which can sometimes separate from the template RNA and re-anneal to another stretch of RNA with a sequence similar to the original model, generating an artefactual chimeric cDNA [25]. Furthermore, the RNA-Seq method used here relies on fragments assembling, which might place false positive on the final sequencing results [26]. This strategy might not be helpful when encountering the needs for *De novo* sequencing without any reference data. Thus, a short sequence of the protein was verified by molecular cloning to prove the accuracy of RNA-Seq. We focused on protein CL4590 due to its high contents, and up to 296 peptides derived from the protein. This specific sequence comparison between the nucleotide sequence of cloned cDNA and the corresponding part in RNA-Seq data showed an extraordinary sequence similarity, which illustrates the result of direct RNA sequencing is reliable.

3.3 Anticancer activity

The total peptides in *Bufo gargarizans* secretion were extracted by water and filtered with 10 KD ultrafiltration. We performed anticancer activity assays on the total peptides and the result of MTT showed obvious inhibition on SMMC-7721 cancer cells in the concentration of 0.1981 μg/ml. Thus, the total peptide extraction obtained from *Venenum Bufonis* are possibly contain peptides that can inhibit the growth of cancer cells (Figure 7).

3.4 Bioinformation analysis

A peptide named Buforin II isolated from the stomach of *Bufo gargarizans* is well known owing to its strong antibacterial and anticancer activity. The cDNA was found by RNA-Seq, but the mature peptide was not found by LC mass spectrometry. It could be speculated that the peptide is not expressed after translation or it only existed in toad stomach [13]. Buforin II shows a complete sequence identity with the N-terminal region of histone 2A. Four peptides identified by LC/MS were found homology to histone H2B (*Gallus gallus*) and one was identified from histone H3 *(Chaperina fusca*) (Supplemental Table 2). Histones belong to alkaline proteins usually existed in eukaryotic cell nucleus who can package and instruct DNA into nucleosomes [27]. Histone H2A and histone H3 mentioned here belong to core histone, which are the principal proteins of chromatin. And all histones have a highly positively charged N-terminus with many lysine and arginine residues [28]. Regularly, antimicrobial peptides are positively charged while the cell membranes have negative charges, they would bind to each other by electrostatic attraction. Thus, five peptides derived from histone proteins are possibly kinds of antimicrobial or anticancer peptides. To sum up, some further analysis need to be done in the future to explore the bioactivity of these peptides.

Apart from these peptides searched in amphibian database with known functions, a large number of peptides obtained by *De novo* analysis without annotation are considered to treasure-houses, expected to be explored in the future. Gene annotation is an ideal method to connect the transcriptome to the peptidome. More importantly, the expression analysis in the transcript and peptide level could describe the network of the overall gene. In the functional analysis of a set of peptides, it is necessary to have a comprehensive analytical tool. Gene ontology is an international standardized gene functional classification system.

The recognized proteins are automatically annotated [29]. The whole transcripts obtained by RNA-Seq was categorized into three main categories: biological process, cellular component and molecular function (Figure 8). Functional annotation and classification provide predicted information that these peptides and proteins most involved in the cell killing, immune process, steroid biosynthesis, antioxidant and antimicrobial activity. Fourteen proteins get from this study were found in the list, four of them were predicted to take part in immune system process, six might participate in the process of metabolic and the rest of them are regarded to have antioxidant activity (Supplemental Table 3), which indicate the directions for following research.

4. Conclusion

We offered extensive information on the parotid secretion of *Bufo gargarizans* transcriptome and peptidome by mass spectrometry and bioinformation analytical methods in this study. Interpretation of the transcriptomic and peptidomics data uncovered a number of candidate peptides with their precursor proteins which involved in the secretion. The peptides also presented promising results of anti-proliferated activity on SMMC-7721. However, the functions of most peptides are unknown, much more work is required to elucidate these proteins and peptides functions, our findings will aid further research aiming to identify key proteins and peptides in the secretion.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The student abroad award from Chinese Scholarship Council. This work was supported by Natural Science Foundation of China (grant number 81673563 and 81102762); Fund of Teaching-Research-Production Cooperation (BY2015008-01); Open Project Program of Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization (ZDXM-1-14, FJGJS-2015-15); A project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD); and Fund of Quality Standardization of Liu-Shen-Wan (BA2016104, ZYBZH-C-JS-30).

Reference

[1] R.J. Boohaker, M.W. Lee, P. Vishnubhotla, L.M. Perez, A.R. Khaled, The use of therapeutic peptides to target and to kill cancer cells. *Curr. med. chem.* 2012, 19, (22), 3794-3804. DOI: 10.2174/092986712801661004.

[2] E. Strandberg, J. Zerweck, D. Horn, G. Pritz, M. Berditsch, P. Wadhwani, A. S. Ulrich, Influence of hydrophobic residues on the activity of the antimicrobial peptide magainin 2 and its synergy with PGLa. *J. Pept. Sci.* 2015, 21, (5), 436-45. DOI: 10.1002/psc.2780.

[3] R. Shams Khozani, D. Shahbazzadeh, N. Harzandi, M. M. Feizabadi, K. Pooshang Bagheri, Kinetics Study of Antimicrobial Peptide, Melittin, in Simultaneous Biofilm Degradation and Eradication of Potent Biofilm Producing MDR Pseudomonas aeruginosa Isolates. *Int. J. Pept. Res. Ther.* 2018. DOI: 10.1007/s10989-018-9675-z.

[4] Y.F. Liu, J. Hu, J.H. Zhang, S.L. Wang, C.F. Wu, Isolation, purification, and N-terminal partial sequence of an antitumor peptide from the venom of the Chinese scorpion Buthus martensii Karsch. *Prep. Biochem. Biotechnol.* 2002, 32, (4), 317-327. DOI:10.1081/PB-120015456.

[5] R.A. Cruciani, J.L. Barker, M. Zasloff, H.C.Chen, O. Colamonici, Antibiotic magainins exert cytolytic activity against transformed cell lines through channel formation. *Proc Natl Acad Sci U S A.* 1991, 88, (9), 3792-3796.

[6] S.D. Gupta, A. Gomes, A. Debnath, A. Saha, A. Gomes, Apoptosis induction in human leukemic cells by a novel protein Bengalin, isolated from Indian black scorpion venom: through mitochondrial pathway and inhibition of heat shock proteins. *Chem.-Biol. Interact.* 2010, 183, (2), 293-303. DOI: 10.1016/j.cbi.2009.11.006.

[7] J. Doyle, C.S. Brinkworth, K.L. Wegener, J.A. Carver, L.E. Llewellyn, I.N. Olver, J.H. Bowie, P.A. Wabnitz, M.J. Tyler, nNOS inhibition, antimicrobial and anticancer activity of the amphibian skin peptide, citropin 1.1 and synthetic modifications. *FEBS. J.* 2003, 270, (6), 1141-1153. DOI: 10.1046/j.1432-1033.2003.03462.x.

[8] F. Gao, X. Wang, Z. Li, A. Zhou, E. Tiffany‐Castiglioni, L. Xie, Y. Qian, Identification of anti‐tumor components from toad venom. *Oncol. Lett.* 2017, 14, (1), 15-22. DoI: 10.3892/ol.2017.6160.

[9] J. Yang, L. Zhu, Z. Wu, Y. Wang, Chinese herbal medicines for induction of remission in advanced or late gastric cancer. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005096.pub4/abstract (accessed on April 2013). DOI: 10.1002/14651858.

[10]S. Wang, X. Zhai, B. Li, Effects of cinobufacini injection on contents of serum thyroid-stimulating hormone and adrenaline in rats. *Journal of Chinese integrative medicine.* 2009, 7, (3), 228-231. DOI: 10.3736/jcim20090306

[11]Y. Liu, Y. Xiao, X. Xue, X. Zhang, X. Liang, Systematic screening and characterization of novel bufadienolides from toad skin using ultra‐performance liquid chromatography/electrospray ionization quadrupole time‐of‐flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2010, 24, (5), 667-678. DOI: 10.1002/rcm.4436.

[12]Y. Su, X. Huang, D. Zhang, Y. Zhang, J. Xie, C. Linh, HPLC separation and determination of bufadienolide in cinobufacini injection. *Chin Tradit Pat Med.* 2003, 25, (1), 24-27.

[13]J.H. Cho, B.H. Sung, S.C. Kim, Buforins: histone H2A-derived antimicrobial peptides from toad stomach. *Biochim Biophys Acta.* 2009, 1788, (8), 1564-1569. DOI: 10.1016/j.bbamem.2008.10.025.

[14]T. Sun, B. Zhan, Y. Gao, A novel cathelicidin from Bufo bufo gargarizans Cantor showed specific activity to its habitat bacteria. *Gene.* 2015, 571, (2), 172-177. DOI: 10.1016/j.gene.2015.06.034.

[15]C.E.H. Schmelzer, R. Schöps, R. Ulbrich-Hofmann, R.H.H. Neubert, K. Raith, Mass spectrometric characterization of peptides derived by peptic cleavage of bovine β-casein. *J. Chromatogr. A.* 2004, 1055, (1), 87-92. DOI: 10.1016/j.chroma.2004.09.003.

[16]I. Vetter, J.L. Davis, L.D. Rash, R. Anangi, M. Mobli, P.F. Alewood, R.J. Lewis, G.F. King, Venomics: a new paradigm for natural products-based drug discovery. *Amino Acids.* 2011, 40, (1), 15-28. DOI: 10.1007/s00726-010-0516-4.

[17]T.F. Abreu, B.N. Sumitomo, N. M. Jr, U.C. Oliveira, G.H. Souza, E.S. Kitano, A. Zelanis, S.M. Serrano, I. Junqueira-De-Azevedo, S.P. Jr, Peptidomics of Acanthoscurria gomesiana spider venom reveals new toxins with potential antimicrobial activity. *J. Proteomics.* 2016, 151, 232. DOI: 10.1016/j.jprot.2016.07.012.

[18]A. Zelanis, A. Keiji Tashima, Unraveling snake venom complexity with ‘omics’ approaches: Challenges and perspectives. *Toxicon.* 2014, 87, 131-134. DOI: 10.1016/j.toxicon.2014.05.011.

[19]P. Escoubas, L. Quinton, G.M. Nicholson, Venomics: Unravelling the complexity of animal venoms with mass spectrometry. *J. Mass Spectrom.* 2008; Vol. 43, p 279-95. DOI: 10.1002/jms.1389.

[20]X. Tang, Y. Zhang, W. Hu, D. Xu, H. Tao, X. Yang, Y. Li, L. Jiang, S. Liang, Molecular diversification of peptide toxins from the tarantula Haplopelma hainanum (Ornithoctonus hainana) venom based on transcriptomic, peptidomic, and genomic analyses. *J. Proteome Res.* 2010, 9, (5), 2550-2564. DOI:10.1021/pr1000016.

[21]J.R. Prashanth, R.J. Lewis, S. Dutertre, Towards an integrated venomics approach for accelerated conopeptide discovery. *Toxicon.* 2012, 60, (4), 470. DOI: 10.1016/j.toxicon.2012.04.340.

[22]S. Dutertre, A.H. Jin, Q. Kaas, A. Jones, P.F. Alewood, R.J. Lewis, Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol Cell Proteomics.* 2013, 12, (2), 312-329. DOI: 10.1074/mcp.M112.021469.

[23]V. Sitprija, Animal toxins and the kidney. *Nat. Clin. Pract. Nephrol.* 2008, 4, (11), 616. DOI: 10.1038/ncpneph0941.

[24]L. Jiang, L. Peng, J. Chen, Y. Zhang, X. Xiong, S. Liang, Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider Ornithoctonus huwena. *Toxicon.* 2008, 51, (8), 1479-1489. DOI: 10.1016/j.toxicon.2008.03.024.

[25]F. Ozsolak, P.M. Milos, Single-molecule direct RNA sequencing without cDNA synthesis. *Wires RNA.* 2011, 2, (4), 565-570. DOI: 10.1002/wrna.84.

[26]F. Ozsolak, P.M. Milos, RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 2011, 12, (2), 87. DOI:10.1038/nrg2934.

[27]R.M. Youngson, Collins dictionary of human biology, Harper Collins Pb., 1St Edition edition, 2006.

[28]A.D.R iggs, S. Bourgeois, M. Cohn, lac repressor—Operator interaction. *J. Mol. Biol.* 1970, 48, (3), 401. DOI: 10.1016/0022-2836(70)90219-6

[29]M.A. Harris, The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004, Vol.32, Database issue, 258-261. DOI: 10.1093/nar/gkh036.

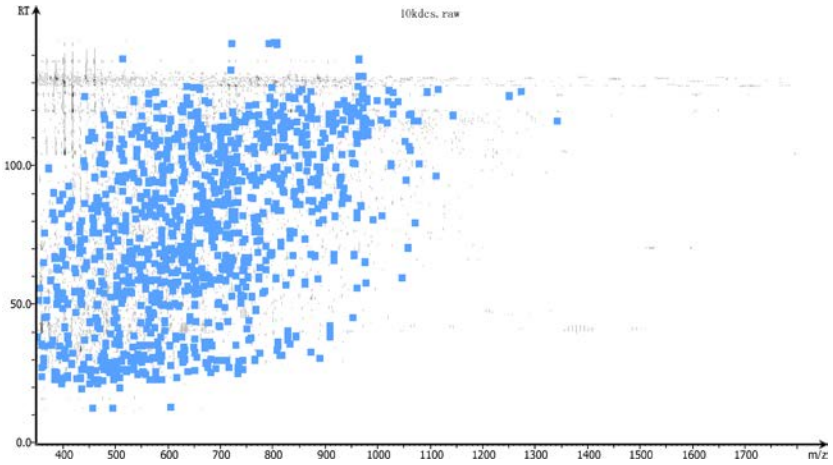


Figure 1. Mass map of *Venenum bufonis* peptides ion data displayed by PEAKS after processing of LTQ-Orbitrap velos Pro.

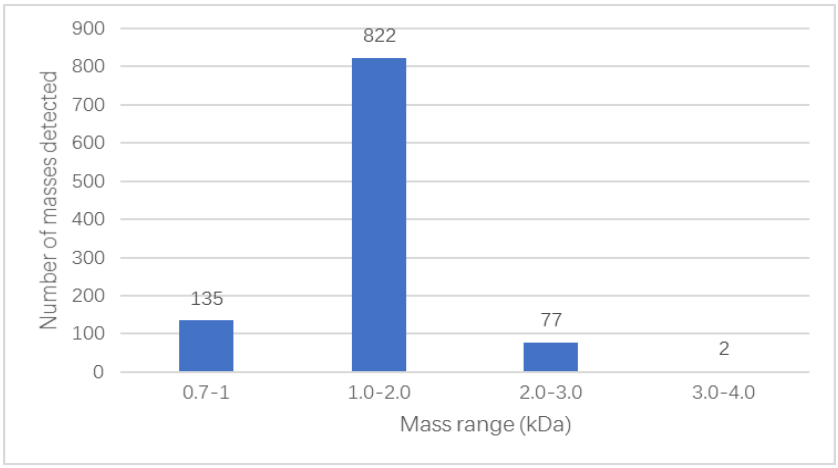
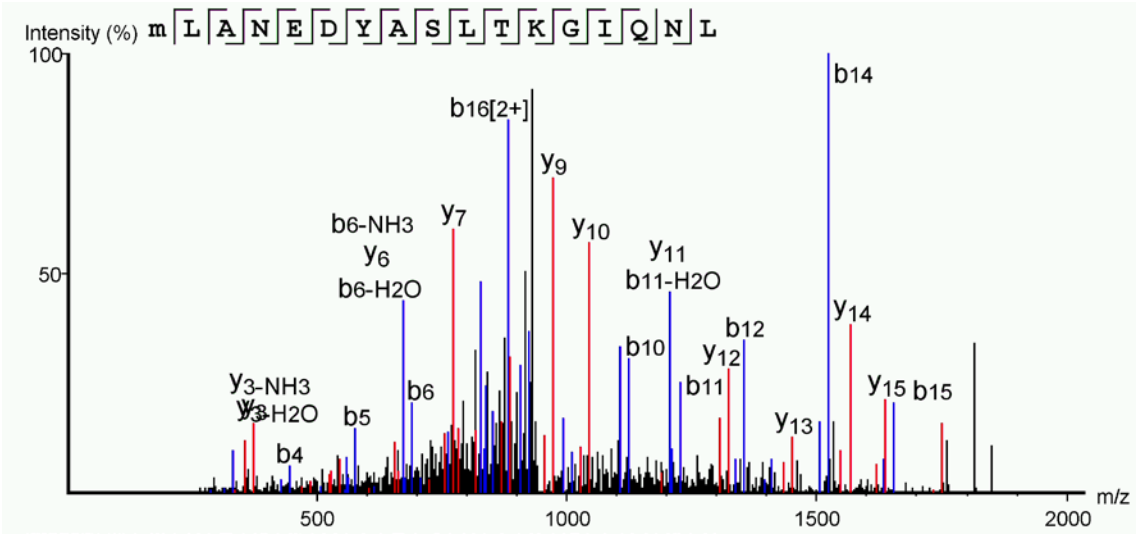
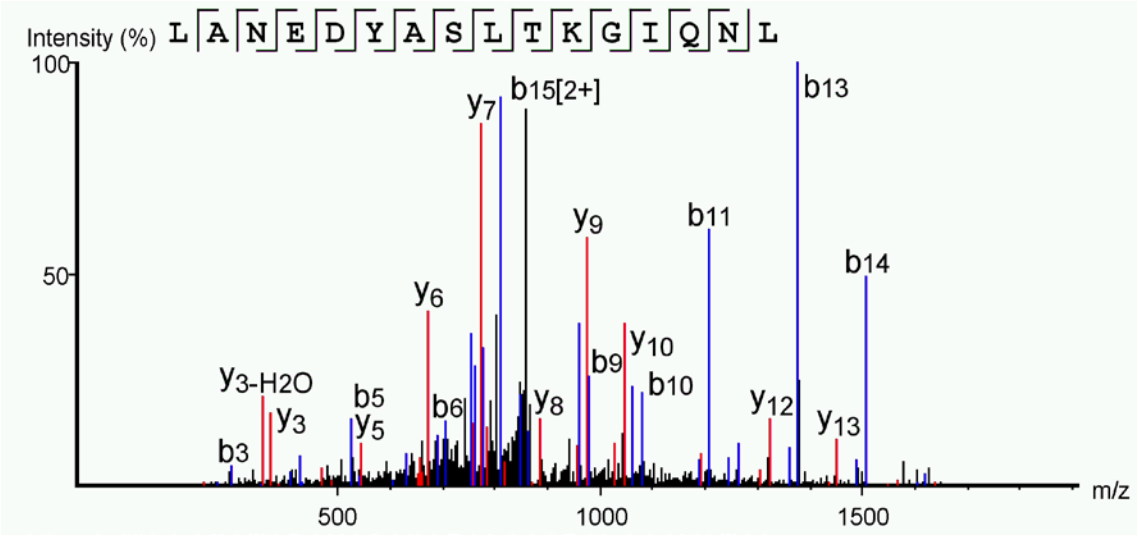


Figure 2. The molecular weight of peptides from *Venenum Bufonis. x-axis* represents the mass ranges of these peptides and y-*axis* represents the number of masses that were detected.



A



B

Figure 3. The MS/MS spectrum of the best unique peptide M(+15.99)LANEDYASLTKGIQNL (m/z 948.9697) (A) and LANEDYASLTKGIQNL (m/z 875.4536) (B) identified in CL4590.

Figure 4. Peptide spectrum matches indicated by blue lines below the sequences of the precusor protein CL4590 (partial).

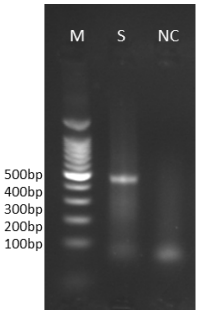


Figure 5. Gel electropherogram of PCR products generated from *Bufo gargarizans* secretion-derived cDNA library. Lane M contains a standard DNA ladder as a marker, and each band represents 100bp increments. Lane S accounts for the PCR product from cDNA library of *Bufo gargarizans* secretion at 62°C. Lane NC represents the PCR product using deionized water at 62°C as a negative control.

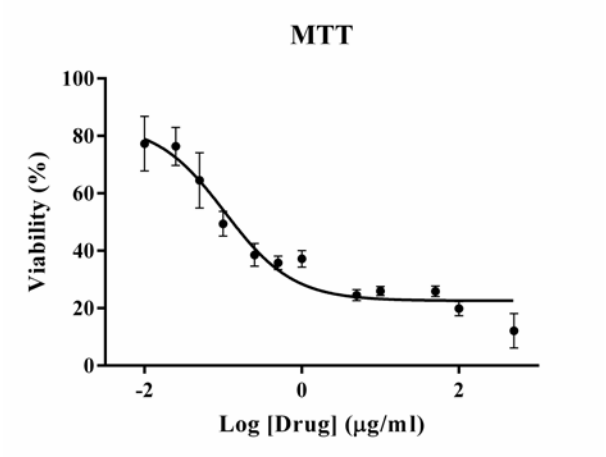
```
CL4590     (365) HSDVPSQYPGNKILNRFGISILERTITNGNYKVPESTTNAYHFLRSVCQFLRDLKNGVEI
Validated    (1) HSDVPSQYPGNKILNRFGISILERTIENGNYKVPESTTNAYHFLRSVCQFLRDLKNGVEI

CL4590     (425) KPPLISWLSQLNQDVSSYLKLPATPLRVSLWQELSYLVQCNLPEVSKENPVKNQSKEAFL
Validated   (61) KPPLISWLSQLNQDVSSYLKLPATPLRVSLWQELSYLVQCNLPEVSKENPVKNQSKEAFL

CL4590     (485) ICLAQEVNCLYDSAQEDTGELDVQGEPTRIDIDGTTLDGDTWRSTGL
Validated  (121) ICLAQEVNCLYDSVQEDTGELDVQGEPTRIDIDGTTLDGDTWRSTGL
```

A

```
CL4590    (1092) CCATTCAGACGTTCCTTCTCAGTATCCAGGAAACAAAATCCTCAATAGATTTGGGATAAG
Validated    (1) CCATTCAGACGTTCCTTCTCAGTATCCAGGAAACAAAATCCTCAATAGATTTGGGATAAG

CL4590    (1152) CATTCTGGAAAGAACCATAACAAATGGCAATTATAAAGTTCCCGAGTCTACTACTAACGC
Validated   (61) CATTCTGGAGAGAACTATAGAAAATGGCAATTATAAAGTTCCCGAGTCTACTACTAACGC

CL4590    (1212) ATATCATTTTCTCAGATCAGTGTGTCAGTTTCTGAGAGACTTGAAAAATGGAGTGGAGAT
Validated  (121) ATATCATTTTCTCAGATCAGTGTGTCAGTTTCTGAGAGACTTGAAAAATGGAGTGGAGAT

CL4590    (1272) TAAACCCCCTCTAATCTCCTGGTTATCACAGCTCAATCAAGATGTCTCTAGCTATCTGAA
Validated  (181) TAAACCCCCTCTAATCTCCTGGTTATCACAGCTCAATCAAGATGTCTCTAGCTATCTGAA

CL4590    (1332) GTTGCCTGCCACACCTCTTAGAGTGTCCCTATGGCAAGAACTTTCATATCTGGTACAATG
Validated  (241) GTTGCCTGCCACACCTCTTAGAGTGTCCCTATGGCAAGAACTTTCATATCTGGTACAATG

CL4590    (1392) TAATTTACCCGAAGTTAGTAAAGAGAATCCAGTGAAAAACCAATCCAAAGAGGCTTTTCT
Validated  (301) TAATTTACCCGAAGTTAGTAAAGAGAATCCAGTGAAAAACCAATCCAAAGAGGCTTTTCT

CL4590    (1452) AATATGTTTGGCTCAAGAAGTCAACTGCCTATATGACTCAGCTCAAGAGGACACAGGAGA
Validated  (361) AATATGTTTGGCTCAAGAAGTCAACTGCCTATATGACTCAGTTCAAGAGGACACAGGAGA

CL4590    (1512) GCTTGATGTTCAAGGAGAACCCACACGTATAGATATTGATGGCACAACTCTTGATGGTGA
Validated  (421) GCTTGATGTTCAAGGAGAACCCACACGTATAGATATTGATGGCACAACTCTTGATGGTGA

CL4590    (1572) TACATGGAGAAGTACAGGGCTA
Validated  (481) TACATGGAGAAGTACAGGGCTA
```

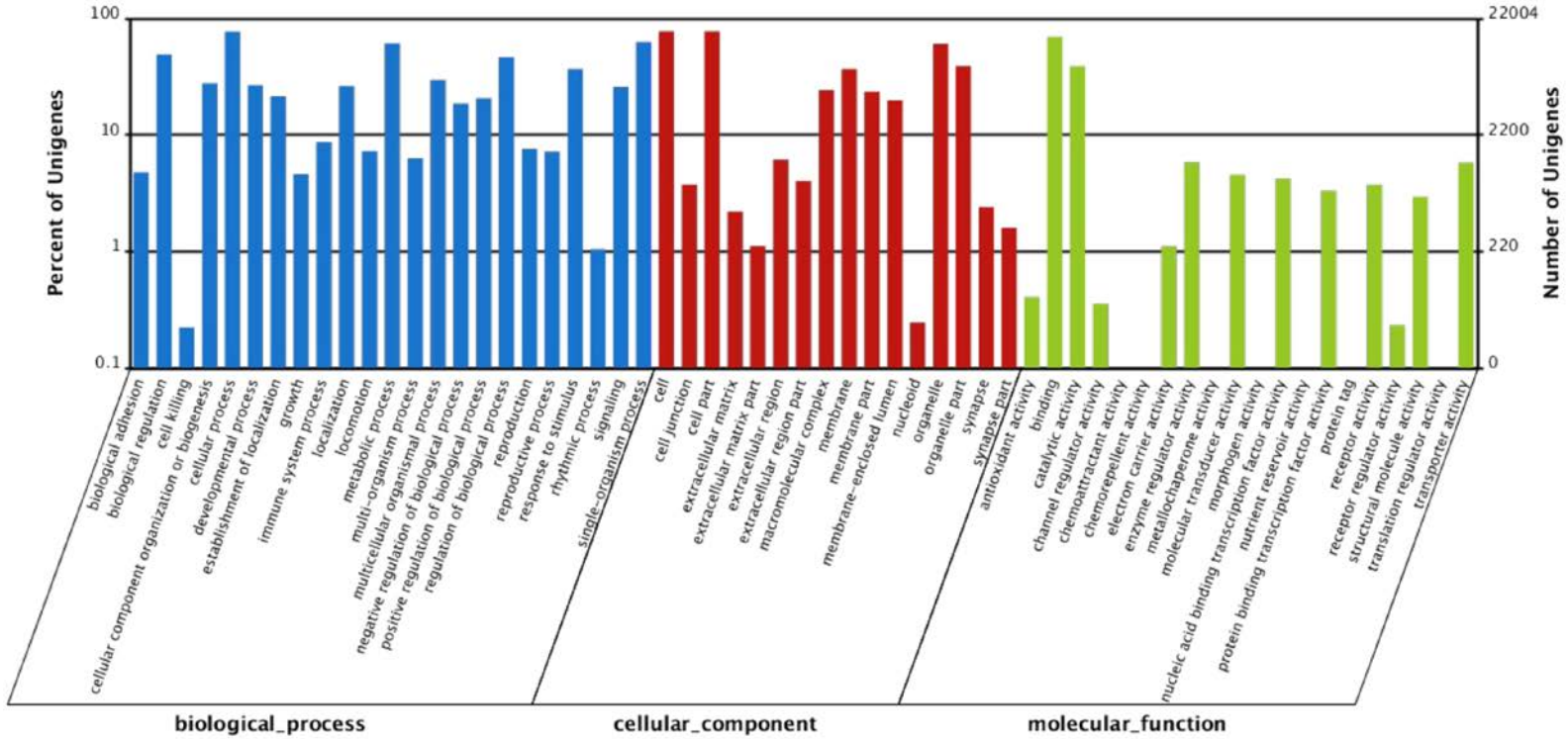
B

Figure 6. (A) Alignment of translated open-reading frame amino acid sequence of the corresponding fragment of CL4590 and cloned fragment(Validated). (B) Alignment of the corresponding fragment of CL4590 and cloned nucleotide sequence (Validated). The discrepant nucleotides or amino acid residues are indicated by blue.



485

486 **Figure 7.** The viability of SMMC-7721 treated by peptides extracted from secretion of
487 *Bufo gargarizans*.



488

489 **Figure 8**. GO classification analysis of Unigenes in All-Unigene. GO functions is showed in *X-axis*. The right *Y-axis* shows the number of genes

490 which have the GO function, and the left *Y-axis* shows the percentage.

Identification of ＜10KD peptides in the water extraction of *Venenum Bufonis* from *Bufo gargarizans* using Nano-LC-MS/MS and *De novo* Sequencing

Yiguo Huo[1,2], Ruoxian Xv[1,], Hongyue Ma[1,*], Jing Zhou[1], Xinping Xi[2,*], Qinan Wu[1], Jinao Duan[1], Mei Zhou[2], Tianbao Chen[2]

[1]Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, and Jiangsu Key Laboratory for High Technology Research of TCM Formulae, College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; [2]Natural Drug Discovery Group, School of Pharmacy, Queen's University, Belfast BT9 7BL, Northern Ireland, UK.

* Correspondence: hongyuema@126.com (H.M.); x.xi@qub.ac.uk (X.X.)

```
    M  L  A  N     E  D  Y     A  S  L     T  K  G  I     Q  N  L     N  F  T
  1 ATGTTGGCTA ATGAAGATTA CGCCTCACTT ACCAAAGGCA TCCAGAATCT AAATTTTACA
    TACAACCGAT TACTTCTAAT GCGGAGTGAA TGGTTTCCGT AGGTCTTAGA TTTAAAATGT
    G  D  S  I     P  C  R     L  L  L     T  G  D  T     A  F  P     V  L  E
 61 GGAGACAGTA TTCCTTGTAG GCTTCTTCTT ACTGGTGACA CAGCATTCCC AGTGCTGGAG
    CCTCTGTCAT AAGGAACATC CGAAGAAGAA TGACCACTGT GTCGTAAGGG TCACGACCTC
    T  P  A  N     N  A  L     I  A  A     S  R  Y  G     K  G  R     L  V  V
121 ACTCCTGCAA ATAATGCTCT CATTGCTGCA TCCAGATACG GCAAGGGCCG CTTGGTTGTC
    TGAGGACGTT TATTACGAGA GTAACGACGT AGGTCTATGC CGTTCCCGGC GAACCAACAG
    M  S  H  E     S  Y  L     N  E  P     Q  F  M  D     L  L  L     N  V  I
181 ATGAGTCATG AATCATATCT GAATGAGCCA CAATTTATGG ATTTACTGCT GAATGTGATA
    TACTCAGTAC TTAGTATAGA CTTACTCGGT GTTAAATACC TAAATGACGA CTTACACTAT
    S  W  L  K     P  S  S     E  A  V     I  G  I  E     N  S  L     G  L  L
241 TCCTGGTTGA AACCATCTTC GGAAGCAGTA ATTGGCATTG AAAACAGCCT AGGTCTCCTG
    AGGACCAACT TTGGTAGAAG CCTTCGTCAT TAACCGTAAC TTTTGTCGGA TCCAGAGGAC
    A  Q  T  L     S  I  S     G  H  K     I  D  T  T     S  G  F     K  E  D
301 GCACAAACTC TTTCTATCTC TGGCCATAAA ATTGACACGA CCTCTGGCTT TAAGGAAGAT
    CGTGTTTGAG AAAGATAGAG ACCGGTATTT TAACTGTGCT GGAGACCGAA ATTCCTTCTA
    L  G  V  F     C  M  T     Q  Y  N     E  S  Q  L     S  E  I     V  S  F
361 TTGGGAGTGT TCTGCATGAC ACAATATAAT GAAAGCCAGT TATCAGAAAT TGTGTCATTT
    AACCCTCACA AGACGTACTG TGTTATATTA CTTTCGGTCA ATAGTCTTTA ACACAGTAAA
    L  R  E  G     G  G  L     L  I  G     A  Q  A  C     N  W  S     E  C  H
421 TTGAGGGAAG GTGGCGGTCT TCTTATTGGA GCTCAAGCCT GCAATTGGTC TGAATGCCAC
    AACTCCCTTC CACCGCCAGA AGAATAACCT CGAGTTCGGA CGTTAACCAG ACTTACGGTG
    P  E  L  N     V  F  F     D  F  P     G  N  K  I     T  S  V     A  G  V
481 CCAGAATTAA ACGTTTTCTT CGATTTTCCT GGAAACAAAA TAACATCTGT AGCTGGAGTC
    GGTCTTAATT TGCAAAAGAA GCTAAAAGGA CCTTTGTTTT ATTGTAGACA TCGACCTCAG
    Y  F  T  N     K  T  G     E  N  G     I  F  P  V     S  E  K     P  P  P
541 TATTTCACTA ATAAGACTGG AGAAAATGGG ATTTTCCCAG TGAGTGAGAA ACCACCACCA
    ATAAAGTGAT TATTCTGACC TCTTTTACCC TAAAAGGGTC ACTCACTCTT TGGTGGTGGT
    F  P  S  F     D  D  V     D  F  S     M  D  L  K     Q  I  F     N  G  V
601 TTTCCGTCTT TTGATGATGT GGACTTCTCC ATGGACCTAA AACAAATCTT CAATGGAGTA
    AAAGGCAGAA AACTACTACA CCTGAAGAGG TACCTGGATT TTGTTTAGAA GTTACCTCAT
    S  S  L  D     I  T  G     S  S  V     V  S  K  L     L  L  H     G  P  L
661 TCCAGTCTGG ATATTACTGG GAGTTCTGTT GTTTCTAAAC TCCTGCTACA CGGACCGTTG
    AGGTCAGACC TATAATGACC CTCAAGACAA CAAAGATTTG AGGACGATGT GCCTGGCAAC
    T  F  P  L     G  L  T     S  S  H     Q  C  F  F     A  A  A     Y  Y  G
721 ACTTTTCCAC TTGGACTGAC AAGCAGCCAC CAGTGTTTTT TTGCTGCTGC ATACTATGGC
    TGAAAAGGTG AACCTGACTG TTCGTCGGTG GTCACAAAAA AACGACGACG TATGATACCG
    R  G  R  V     V  V  G     S  H  E     G  Y  I  T     R  L  E     L  K  S
781 AGAGGACGTG TTGTAGTGGG GTCTCATGAA GGCTATATTA CAAGACTAGA GCTGAAGTCT
    TCTCCTGCAC AACATCACCC CAGAGTACTT CCGATATAAT GTTCTGATCT CGACTTCAGA
    F  I  L  N     A  I  S     W  L  D     A  G  R  S     G  R  I     G  V  N
841 TTTATCCTAA ATGCCATTTC ATGGCTGGAT GCAGGAAGAA GTGGGAGGAT TGGTGTTAAT
    AAATAGGATT TACGGTAAAG TACCGACCTA CGTCCTTCTT CACCCTCCTA ACCACAATTA
    K  A  M  R     G  L  I     P  I  L     Q  A  E  G     I  S  C     A  M  S
901 AAGGCAATGC GAGGCCTGAT TCCCATACTG CAGGCTGAAG GTATCTCTTG TGCAATGTCC
    TTCCGTTACG CTCCGGACTA AGGGTATGAC GTCCGACTTC CATAGAGAAC ACGTTACAGG
    N  V  S  S     E  F  S     I  Y  C     C  N  S  Y     S  D  A     E  V  D
961 AACGTGAGTT CTGAATTCAG CATTTATTGT TGCAATTCTT ATAGTGATGC AGAAGTAGAC
    TTGCACTCAA GACTTAAGTC GTAAATAACA ACGTTAAGAA TATCACTACG TCTTCATCTG
    K  I  H  Q     F  V  A     E  G  G     S  L  L  I     A  G  H     A  W  Y
1021 AAAATCCATC AGTTTGTTGC AGAAGGAGGA AGTCTTCTTA TTGCTGGCCA TGCCTGGTAC
    TTTTAGGTAG TCAAACAACG TCTTCCTCCT TCAGAAGAAT AACGACCGGT ACGGACCATG
    W  S  Y  S     H  S  D     V  P  S     Q  Y  P  G     N  K  I     L  N  R
1081 TGGTCATACT CCCATTCAGA CGTTCCTTCT CAGTATCCAG GAAACAAAAT CCTCAATAGA
    ACCAGTATGA GGGTAAGTCT GCAAGGAAGA GTCATAGGTC CTTTGTTTTA GGAGTTATCT
    F  G  I  S     I  L  E     R  T  I     T  N  G  N     Y  K  V     P  E  S
1141 TTTGGGATAA GCATTCTGGA AAGAACCATA ACAAATGGCA ATTATAAAGT TCCCGAGTCT
    AAACCCTATT CGTAAGACCT TTCTTGGTAT TGTTTACCGT TAATATTTCA AGGGCTCAGA
    T  T  N  A     Y  H  F     L  R  S     V  C  Q  F     L  R  D     L  K  N
1201 ACTACTAACG CATATCATTT TCTCAGATCA GTGTGTCAGT TTCTGAGAGA CTTGAAAAAT
    TGATGATTGC GTATAGTAAA AGAGTCTAGT CACACAGTCA AAGACTCTCT GAACTTTTTA
```

```
      G  V  E  I   K  P  P   L  I  S   W  L  S  Q   L  N  Q   D  V  S
1261  GGAGTGGAGA TTAAACCCCC TCTAATCTCC TGGTTATCAC AGCTCAATCA AGATGTCTCT
      CCTCACCTCT AATTTGGGGG AGATTAGAGG ACCAATAGTG TCGAGTTAGT TCTACAGAGA
      S  Y  L  K   L  P  A   T  P  L   R  V  S  L   W  Q  E   L  S  Y
1321  AGCTATCTGA AGTTGCCTGC CACACCTCTT AGAGTGTCCC TATGGCAAGA ACTTTCATAT
      TCGATAGACT TCAACGGACG GTGTGGAGAA TCTCACAGGG ATACCGTTCT TGAAAGTATA
      L  V  Q  C   N  L  P   E  V  S   K  E  N  P   V  K  N   Q  S  K
1381  CTGGTACAAT GTAATTTACC CGAAGTTAGT AAAGAGAATC CAGTGAAAAA CCAATCCAAA
      GACCATGTTA CATTAAATGG GCTTCAATCA TTTCTCTTAG GTCACTTTTT GGTTAGGTTT
      E  A  F  L   I  C  L   A  Q  E   V  N  C  L   Y  D  S   A  Q  E
1441  GAGGCTTTTC TAATATGTTT GGCTCAAGAA GTCAACTGCC TATATGACTC AGCTCAAGAG
      CTCCGAAAAG ATTATACAAA CCGAGTTCTT CAGTTGACGG ATATACTGAG TCGAGTTCTC
      D  T  G  E   L  D  V   Q  G  E   P  T  R  I   D  I  D   G  T  T
1501  GACACAGGAG AGCTTGATGT TCAAGGAGAA CCCACACGTA TAGATATTGA TGGCACAACT
      CTGTGTCCTC TCGAACTACA AGTTCCTCTT GGGTGTGCAT ATCTATAACT ACCGTGTTGA
      L  D  G  D   T  W  R   S  T  G   L  Y  L  S   P  N  K   K  A  T
1561  CTTGATGGTG ATACATGGAG AAGTACAGGG CTATACTTGT CTCCTAACAA AAAAGCAACA
      GAACTACCAC TATGTACCTC TTCATGTCCC GATATGAACA GAGGATTGTT TTTTCGTTGT
      L  V  F  P   A  S  A   A  G  K   G  L  Q  V   Q  V  G   C  Q  S
1621  CTTGTATTTC CAGCCTCAGC TGCAGGGAAG GGTCTCCAGG TACAAGTTGG CTGTCAATCC
      GAACATAAAG GTCGGAGTCG ACGTCCCTTC CCAGAGGTCC ATGTTCAACC GACAGTTAGG
      D  D  L  S   G  A  A   E  L  C   R  A  P  V   V  I  H   R  K  N
1681  GATGACCTGA GCGGAGCAGC AGAGTTGTGT CGTGCTCCAG TGGTTATACA TAGAAAGAAC
      CTACTGGACT CGCCTCGTCG TCTCAACACA GCACGAGGTC ACCAATATGT ATCTTTCTTG
      V  V  D  E   K  V  V   I  S  C   I  W  G  G   L  L  Y   V  I  V
1741  GTTGTTGATG AGAAGGTTGT GATATCCTGT ATCTGGGGAG GACTTCTTTA CGTTATTGTA
      CAACAACTAC TCTTCCAACA CTATAGGACA TAGACCCCTC CTGAAGAAAT GCAATAACAT
      K  G  K  S   Q  L  G   N  V  P   V  T  I  Y   G  A  E   P  A  P
1801  AAGGGGAAGA GTCAACTAGG AAATGTTCCA GTTACAATCT ATGGGGCAGA ACCAGCTCCA
      TTCCCCTTCT CAGTTGATCC TTTACAAGGT CAATGTTAGA TACCCCGTCT TGGTCGAGGT
      T  F  I  N   G  Q  T   K  V  S   S  W  L  D   T  L  R   T  S  T
1861  ACATTTATAA ATGGACAAAC TAAAGTTTCT TCGTGGTTGG ATACATTACG CACTTCAACA
      TGTAAATATT TACCTGTTTG ATTTCAAAGA AGCACCAACC TATGTAATGC GTGAAGTTGT
      S  P  W  A   E  L  S   A  E  N   I  T  L  T   V  P  T   D  K  V
1921  TCTCCTTGGG CTGAGCTTAG CGCAGAGAAT ATTACCCTGA CTGTCCCTAC TGATAAAGTC
      AGAGGAACCC GACTCGAATC GCGTCTCTTA TAATGGGACT GACAGGGATG ACTATTTCAG
      R  S  L  E   D  P  E   S  L  M   S  L  W  D   N  I  M   S  T  V
1981  CGCTCACTAG AGGATCCTGA ATCGCTTATG TCTTTATGGG ATAACATCAT GTCAACAGTG
      GCGAGTGATC TCCTAGGACT TAGCGAATAC AGAAATACCC TATTGTAGTA CAGTTGTCAC
      T  D  L  A   S  I  P   K  K  L   P  R  P  E   R  I  V   A  D  V
2041  ACTGATCTGG CATCTATCCC AAAGAAACTG CCTCGTCCAG AAAGAATTGT GGCCGATGTT
      TGACTAGACC GTAGATAGGG TTTCTTTGAC GGAGCAGGTC TTTCTTAACA CCGGCTACAA
      Q  I  S  A   G  F  M   H  A  G   Y  P  I  M   C  H  V   P  T  A
2101  CAGATCTCAG CAGGTTTTAT GCATGCTGGA TACCCAATAA TGTGCCATGT GCCAACAGCA
      GTCTAGAGTC GTCCAAAATA CGTACGACCT ATGGGTTATT ACACGGTACA CGGTTGTCGT
      S  S  L  V   S  V  T   S  L  K   K  G  M  W   G  A  A   H  E  L
2161  TCTTCTTTAG TAAGTGTTAC ATCATTGAAG AAAGGCATGT GGGGGGCTGC ACATGAGCTT
      AGAAGAAATC ATTCACAATG TAGTAACTTC TTTCCGTACA CCCCCCGACG TGTACTCGAA
      G  H  N  Q   Q  R  G   V  W  E   F  P  P  H   T  T  E   A  T  C
2221  GGTCACAACC AGCAGAGAGG AGTTTGGGAA TTCCCTCCTC ACACTACAGA AGCCACCTGT
      CCAGTGTTGG TCGTCTCTCC TCAAACCCTT AAGGGAGGAG TGTGATGTCT TCGGTGGACA
      N  L  W  S   V  Y  V   H  E  I   L  L  G  I   P  R  D   N  A  H
2281  AACCTGTGGT CAGTATATGT ACATGAAATC TTGCTGGGGA TCCCAAGAGA TAACGCTCAT
      TTGGACACCA GTCATATACA TGTACTTTAG AACGACCCCT AGGGTTCTCT ATTGCGAGTA
      P  A  L  K   P  G  D   R  E  K   R  I  K  Q   Y  L  K   N  G  A
2341  CCTGCTCTCA AGCCAGGAGA CAGAGAAAAA AGAATCAAGC AGTATCTGAA GAATGGAGCC
      GGACGAGAGT TCGGTCCTCT GTCTCTTTTT TCTTAGTTCG TCATAGACTT CTTACCTCGG
      K  L  D  E   W  S  V   W  T  A   L  E  T  Y   L  Q  L   Q  E  C
2401  AAACTTGATG AGTGGAGCGT GTGGACAGCG CTTGAGACTT ATCTTCAGCT ACAAGAGTGC
      TTTGAACTAC TCACCTCGCA CACCTGTCGC GAACTCTGAA TAGAAGTCGA TGTTCTCACG
      F  G  W  D   P  F  K   Q  V  F   S  E  Y  Q   T  M  S   N  V  S
2461  TTTGGCTGGG ATCCCTTCAA GCAAGTTTTC TCAGAGTATC AGACCATGTC CAATGTCAGT
      AAACCGACCC TAGGGAAGTT CGTTCAAAAG AGTCTCATAG TCTGGTACAG GTTACAGTCA
```

```
      N  V  K  D   V  K  M   N  L  W   A  E  K  F   S  H  A   V  N  K
2521 AATGTCAAGG ATGTTAAGAT GAATCTCTGG GCAGAGAAGT TTTCTCACGC AGTCAATAAG
     TTACAGTTCC TACAATTCTA CTTAGAGACC CGTCTCTTCA AAAGAGTGCG TCAGTTATTC
      N  L  A  S   F  F  K   A  W  G   W  P  I  D   D  A  T   T  E  K
2581 AACCTAGCTT CATTCTTCAA GGCTTGGGGA TGGCCCATTG ATGATGCTAC CACCGAAAAA
     TTGGATCGAA GTAAGAAGTT CCGAACCCCT ACCGGGTAAC TACTACGATG GTGGCTTTTT
      L  S  I  L   S  E  W   E  E  N   P  M  N  K   Y  V  S   *
2641 TTATCCATAT TAAGTGAATG GGAAGAGAAT CCAATGAATA AATATGTCAG CTAAATTTCA
     AATAGGTATA ATTCACTTAC CCTTCTCTTA GGTTACTTAT TTATACAGTC GATTTAAAGT
2701 TAAAAGAAAT ATATAAAATG ATCCTTCTAT AGTAATATGC TTTTGGTGGC TTAAGAGGAA
     ATTTTCTTTA TATATTTTAC TAGGAAGATA TCATTATACG AAAACCACCG AATTCTCCTT
2761 TGCATTATAG GCTCAAAAGT GTTAAAGGAA AATATAAAAA TCACATCATA ATAATATTGC
     ACGTAATATC CGAGTTTTCA CAATTTCCTT TTATATTTTT AGTGTAGTAT TATTATAACG
2821 AACAGAGCCT GTTAGATTAT CATCTGAATA TGGAAAACTA ATAGAATATA ATCATATGCA
     TTGTCTCGGA CAATCTAATA GTAGACTTAT ACCTTTTGAT TATCTTATAT TAGTATACGT
2881 ATGCATGCAA TGTTCAATAA AGAATTCAGA TTCCTGCTCC AACTTAGTTT TGTTTTTTTG
     TACGTACGTT ACAAGTTATT TCTTAAGTCT AAGGACGAGG TTGAATCAAA ACAAAAAAAC
2941 TTTTTTT
     AAAAAAA
```

Figure S1. The transcript obtained from the de novo assembled clean reads and the translated amino acid sequence is shown above the nucleotides. The stop codon is indicated by an asterisk.


Acetylation (Protein N-term) (+42.01)
Amidation (-0.98)
Carboxymethyl (KW, HNW-term) (+58.01)
Deamidation (NQ) (+0.98)
Dehydration (-18.01)
Ethylation (+28.03)
Formylation (+27.99)
Levuglandinyl-lysine anhydrolactam adduct (+314.19)
Methylmalonylation on Serine (+100.02)
Oxidation (M) (+15.99)
Oxidation (MW) (+15.99)
Phosphorylation (STY) (+79.97)
Pyro-glu from Q (-17.03)
Ubiquitin (+114.04)
Acetylation (Protein N-term) (+42.01), Oxidation (M) (+15.99)
D->N
N->D
1 MLANEDYASL TKGIQNLNFT GDSIPCRLLL TGDTAFPVLE TPANNALIAA SRYGKGRLVV MSHESYLNEP QFMDLLLNVI
81 SWLKPSSEAV IGIENSLGLL AQTLSISGHK IDTTSGFKED LGVFCMTQYN ESQLSEIVSF LREGGGLLIG AQACNWSECH
161 PELNVFFDFP GNKITSVAGV YFTNKTGENG IFPVSEKPPP FPSFDDVDFS MDLKQIFNGV SSLDITGSSV VSKLLLHGPL
241 TFPLGLTSSH QCFFAAAYYG RGRVVVGSHE GYITRLELKS FILNAISWLD AGRSGRIGVN KAMRGLIPIL QAEGISCAMS
321 NVSSEFSIYC CNSYSDAEVD KIHQFVAEGG SLLIAGHAWY WSYSHSDVPS QYPGNKILNR FGISILERTI TNGNYKVPES
401 TTNAYHFLRS VCQFLRDLKN GVEIKPPLIS WLSQLNQDVS SYLKLPATPL RVSLWQELSY LVQCNLPEVS KENPVKHQSK
481 EAFLICLAQE VNCLYDSAQE DTGELDVQGE PTRIDIDGTT LDGDTWRSTG LYLSPNKKAT LVFPASAAGK GLQVQVGCQS

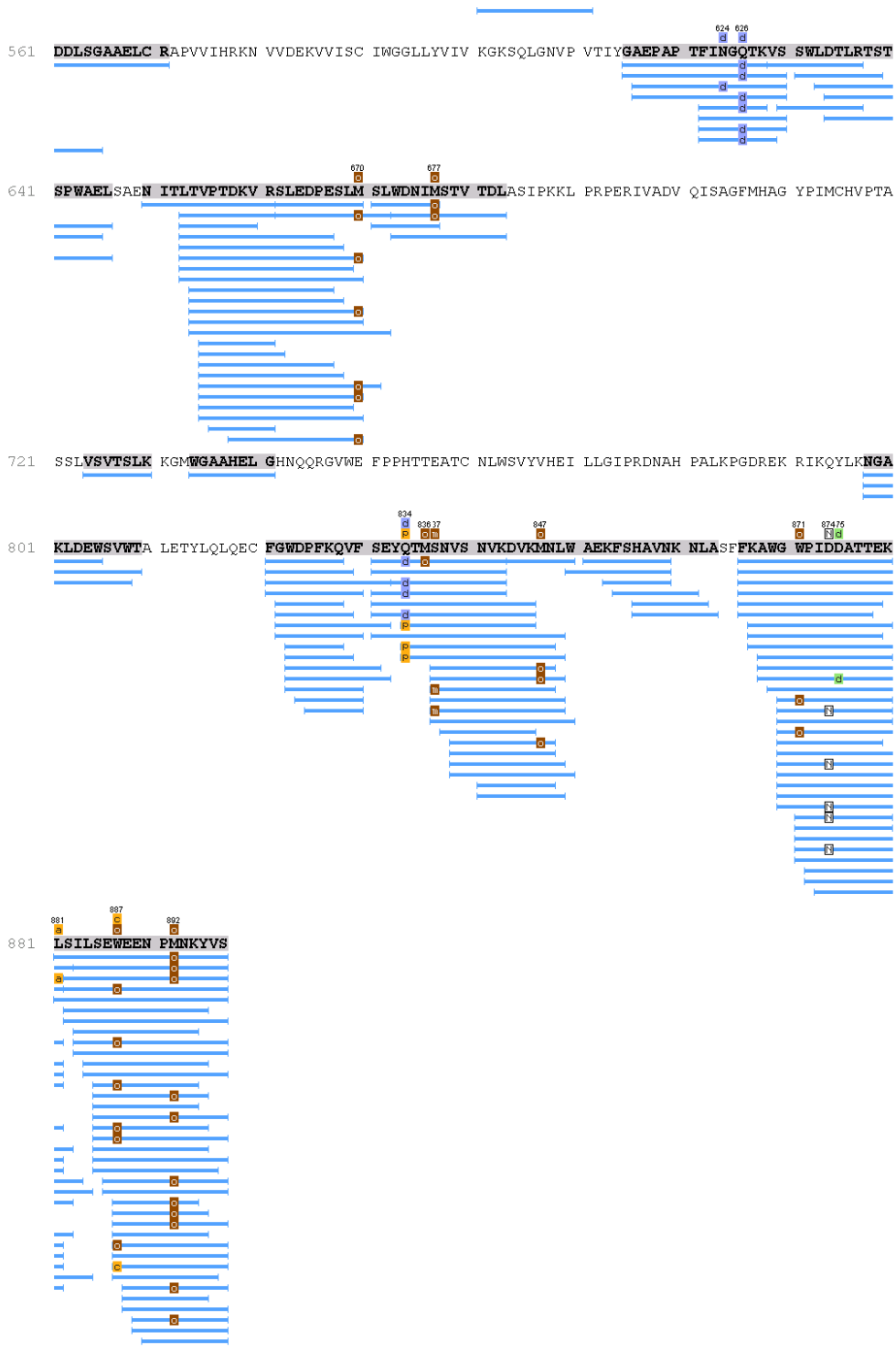


Figure S2. Peptide spectrum matches indicated by blue lines below the sequences of the protein identified.

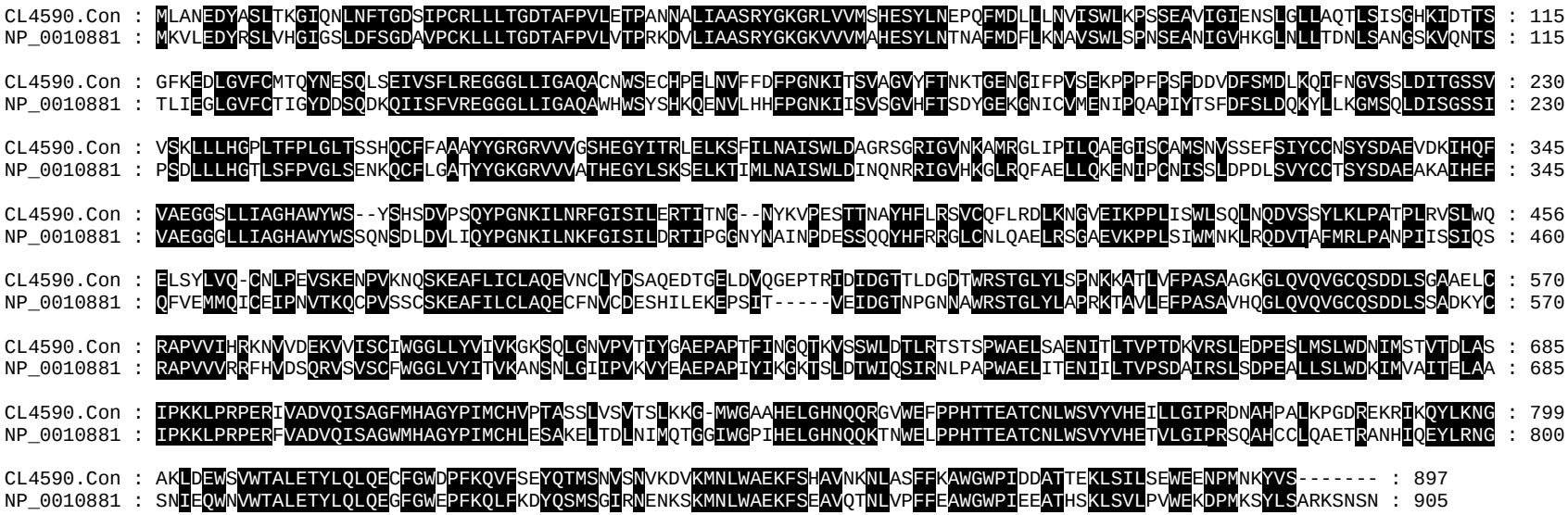


Figure S3. The amino acids sequence alignment of protein CL4590 and NP_0010881.

Supplement Table 1

Protein Accession	Peptide	Unique	-10lgP	Mass	Length	ppm	m/z	z	RT	Intensity	Scan	PTM	AScore
CL2275.Contig5_All	F.TVNTKDAGEGGLS.L	Y	88.36	1247.6	13	0	624.807	2	38.51	9.92E+05	5944		
CL2600.Contig1_All	F.MVDYGTEPDGPVLAH.E	Y	91.84	1599.72	15	0.1	800.869	2	85.11	3.83E+06	18737		
CL4590.Contig1_All	M(+15.99)LANEDYASLTKGIQNL.N	Y	123.75	1895.93	17	-2.7	948.97	2	117.67	6.07E+05	28435	Oxidation	M1:Oxidation (M):1000.00
CL4590.Contig1_All	D.IDGTTLDGDTWRSTGL.Y	Y	113.89	1706.81	16	1	854.414	2	111.39	1.03E+06	26449		
CL4590.Contig1_All	MLANEDYASLTKGIQNL.N	Y	110.97	1879.93	17	1.4	940.976	2	119.5	1.42E+06	29008		
CL4590.Contig1_All	M.LANEDYASLTKGIQNL.N	Y	110.48	1748.89	16	-1	875.454	2	117.04	1.09E+06	28227		
CL4590.Contig1_All	R.IDIDGTTLDGDTWR.S	Y	109.34	1576.74	14	1.2	789.377	2	111.23	1.92E+06	26402		
CL4590.Contig1_All	L.Y(+28.03)DSAQEDTGELDVQGEPTR.I	Y	109.19	2136.94	19	8.2	1069.49	2	79.56	5.71E+04	17114	Ethylation	Y1:Ethylation:10.11
CL4590.Contig1_All	Q.GEPTRIDIDGTTLDGD(-18.01)TW.R	Y	108.82	1942.89	18	2.1	972.455	2	117.32	6.09E+05	28318	Dehydrati	D16:Dehydration:8.22
CL4590.Contig1_All	L.ANEDYASLTKGIQNL.N	Y	108.77	1635.81	15	-0.3	818.912	2	114.87	2.43E+07	27520		
CL4590.Contig1_All	A.NEDYASLTKGIQNL.N	Y	105.14	1564.77	14	1.1	783.395	2	114.27	9.05E+06	27334		
CL4590.Contig1_All	A.NEDYASLTKGIQN.L	Y	104.23	1451.69	13	1	726.853	2	80.86	9.61E+05	17488		
CL4590.Contig1_All	L.YDSAQEDTGELDVQGEPTR.I	Y	103.39	2108.91	19	-5.2	1055.46	2	70.48	1.73E+05	14412		
CL4590.Contig1_All	L.ANEDYASLTKGIQN.L	Y	102.43	1522.73	14	-0.4	762.37	2	81.34	1.73E+06	17630		
CL4590.Contig1_All	M.SHESYLNEPQFM(+15.99)DL.L	Y	102.02	1724.74	14	6.9	863.381	2	111.67	3.03E+05	26533	Oxidation	M12:Oxidation (M):1000.00
CL4590.Contig1_All	T.VPTDKVRSLEDPESL.M	Y	101.6	1683.87	15	1.1	842.942	2	86.85	1.70E+06	19247		
CL4590.Contig1_All	D.IDGTTLD(-18.01)GDTWRSTGL.Y	Y	101.28	1688.8	16	0.3	845.408	2	112.46	1.57E+06	26770	Dehydrati	D7:Dehydration:21.74
CL4590.Contig1_All	F.KAWGWPIDDATTEKL.S	Y	100.95	1729.87	15	-1.9	865.939	2	117.49	2.28E+06	28377		
CL4590.Contig1_All	R.IDIDGTT(-18.01)LDGDTWR.S	Y	100.9	1558.73	14	-0.6	780.37	2	114.17	6.91E+05	27300	Dehydrati	T7:Dehydration:0.00
CL4590.Contig1_All	D.TGELDVQGEPTRIDIDGTTLDGD(-18.01)TW	Y	100.27	2841.34	26	1.4	948.122	3	116.18	2.28E+05	27948	Dehydrati	D23:Dehydration:0.00
CL4590.Contig1_All	Y.DSAQEDTGELDVQGEPTR.I	Y	100.02	1945.85	18	5	973.937	2	65.35	3.11E+04	12931		
CL4590.Contig1_All	G.WPIDDATTEKL.S	Y	99.77	1287.63	11	1	644.825	2	106.43	1.42E+06	24979		
CL4590.Contig1_All	F.FKAWGWPIDDATTEKL.S	Y	99.21	1876.94	16	-5.6	939.47	2	121.27	1.11E+07	29469		
CL4590.Contig1_All	L.ANEDYASLTKGIQNLN.F	Y	98.17	1749.85	16	-0.8	875.933	2	104.34	2.65E+05	24431		
CL4590.Contig1_All	W.GWPIDDATTEKL.S	Y	97.75	1344.66	12	0.3	673.336	2	110.63	1.54E+07	26224		
CL4590.Contig1_All	M(+42.01)LANEDYASLTKGIQNL.N	Y	97.19	1921.95	17	1.3	961.981	2	128.76	1.76E+06	31450	Acetylatio	M1:Acetylation (Protein N-term):1000.00
CL4590.Contig1_All	Q.GEPTRIDIDGTTLDGDTWR.S	Y	97.17	2117	19	-0.4	1059.51	2	105.79	9.96E+06	24795		
CL4590.Contig1_All	M(+42.01)(+15.99)LANEDYASLTKGIQNL.N	Y	97.15	1937.94	17	1.4	969.979	2	125.05	1.05E+07	30561	Acetylatio	M1:Acetylation (Protein N-term):1000.00;M1:Oxidation (M):1000.00
CL4590.Contig1_All	D.IDGTTLDGDTWR.S	Y	96.21	1348.63	12	0.8	675.321	2	85.99	5.79E+06	18991		
CL4590.Contig1_All	D.IDGT(-18.01)TLDGDTWRSTGL.Y	Y	96.2	1688.8	16	1.9	845.409	2	113.58	0	27110	Dehydrati	T4:Dehydration:5.47
CL4590.Contig1_All	D.IDGT(-18.01)TLDGDTWR.S	Y	95.65	1330.62	12	0.2	666.315	2	90.81	3.89E+06	20414	Dehydrati	T4:Dehydration:15.91
CL4590.Contig1_All	M(+42.01)(+15.99)LANEDYASLTK.G	Y	95.25	1412.65	12	-0.2	707.332	2	96.66	1.16E+06	22142	Acetylatio	M1:Acetylation (Protein N-term):1000.00;M1:Oxidation (M):1000.00
CL4590.Contig1_All	T.LVFPASAAGKGLQ.V	Y	94.96	1257.71	13	-0.4	629.861	2	96.18	1.00E+06	21999		
CL4590.Contig1_All	T.VPTDKVRSLEDPESLM(+15.99).S	Y	94.21	1830.9	16	-6.3	916.453	2	81.86	3.33E+05	17785	Oxidation	M16:Oxidation (M):1000.00
CL4590.Contig1_All	K.AWGWPIDDATTEKL.S	Y	94.18	1601.77	14	0.3	801.894	2	122.79	2.42E+06	29907		
CL4590.Contig1_All	T.VPTDKVRSLEDPESLM(+15.99)S.L	Y	94.15	1917.94	17	4.4	959.979	2	81.25	2.89E+04	17604	Oxidation	M16:Oxidation (M):1000.00
CL4590.Contig1_All	W.GWPIDDATTEKLS.I	Y	93.26	1431.69	13	0.6	716.852	2	102.76	2.64E+06	23967		
CL4590.Contig1_All	T.LTVPTDKVRSLEDPESL.M	Y	91.91	1898	17	0.3	950.007	2	104.57	5.02E+05	24498		
CL4590.Contig1_All	M(+42.01)(+15.99)LANEDYASLTKGIQN.L	Y	91.23	1824.86	16	-0.3	913.435	2	112.83	7.62E+05	26881	Acetylatio	M1:Acetylation (Protein N-term):1000.00;M1:Oxidation (M):1000.00
CL4590.Contig1_All	V.FPASAAGKGLQ.V	Y	90.66	1045.56	11	0.7	523.785	2	59.63	3.76E+06	11253		
CL4590.Contig1_All	N.EDYASLTKGIQNL.N	Y	89.94	1450.73	13	0.9	726.373	2	114.31	5.90E+06	27347		
CL4590.Contig1_All	L.SEWEENPM(+15.99)NKY.V	Y	89.93	1441.58	11	-1	721.798	2	66.91	3.05E+06	13364	Oxidation	M8:Oxidation (M):1000.00
CL4590.Contig1_All	F.SEYQ(+.98)TMSNVSNVKDVK.M	Y	89.7	1828.85	16	-3.4	915.43	2	68.19	7.55E+05	13731	Deamidati	Q4:Deamidation (NQ):48.12
CL4590.Contig1_All	R.IDIDGT(-18.01)TLDGDTWR.S	Y	89.67	1558.73	14	1.1	780.371	2	114.47	6.91E+05	27394	Dehydrati	T6:Dehydration:7.32
CL4590.Contig1_All	R.IDIDGTTLDGD(-18.01)TWR.S	Y	89.65	1558.73	14	0.8	780.371	2	115.28	3.30E+06	27656	Dehydrati	D11:Dehydration:0.00
CL4590.Contig1_All	V.ISWLKPSSEAVIGIEN.S	Y	89.42	1741.93	16	-0.6	871.969	2	117.43	1.19E+06	28353		
CL4590.Contig1_All	Q.GEPTRIDIDGTTLDGDTWRSTGL.Y	Y	89.38	2475.19	23	2.5	826.072	3	116.51	1.94E+06	28058		
CL4590.Contig1_All	F.FDFPGNKITSV.A	Y	89.02	1223.62	11	0.5	612.817	2	110.1	7.05E+05	26064		
CL4590.Contig1_All	D.YASLTKGIQNL.N	Y	88.95	1206.66	11	0.3	604.338	2	99.69	3.70E+06	23054		
CL4590.Contig1_All	L.SEWEENPM(+15.99)NKYVS	Y	88.84	1627.68	13	0.7	814.849	2	72.25	4.31E+07	14930	Oxidation	M8:Oxidation (M):1000.00

Supplement Table 2

proteins identified by homology search with PEAKS of peptide sequences obtained by de novo analysis of *B. gargarizans* secretion. Database: Uniprot (amphibian).

Peptide	-10lgP	Mass	m/z	RT	Intensity	Accession	Protein Description
HAVSEGTKAVT	73.53	1098.5669	550.291	26	2.67E+05	P0C1H3|H2B1_CHICK	Histone H2B 1/2/3/4/6 (Gallus gallus)
HAVSEGTKAVTK	68.14	1226.6619	409.8929	22.72	3.22E+05	P0C1H3|H2B1_CHICK	
IFERIAGEASRLA	59.55	1431.7833	716.8997	106.22	1.84E+05	P0C1H3|H2B1_CHICK	
PEPAKSAPAPK	49.85	1091.5974	546.8065	22.99	6.16E+04	P0C1H3|H2B1_CHICK	
DSYVGDEAE(sub Q)SKR	48.54	1354.6001	678.3071	36.73	2.00E+04	P53506|ACT8_XENLA	Actin cytoplasmic type 8 (Xenopus laevis)
EAPLNPKAN	51.69	952.4977	477.256	32.37	1.87E+05	P53506|ACT8_XENLA	
GQKDSYVGDEAQ	68.71	1295.563	648.7886	34.87	3.11E+04	P53506|ACT8_XENLA	
ISKQEYDESGPS	78.92	1338.5939	670.3048	37.4	4.43E+05	P53506|ACT8_XENLA	
LVV(sub I)DNGSGMCK	48.2	1121.5209	561.7679	58.35	3.97E+05	P53506|ACT8_XENLA	
MGQKDSYVGDEAQ	60.93	1426.6034	714.3116	43.9	1.10E+05	P53506|ACT8_XENLA	
MGQKDSYVGDEAQSKRGILT	73.42	2182.0688	728.3621	68	8.96E+05	P53506|ACT8_XENLA	
RVAPEEHPV	59.39	1032.5352	517.2799	33	6.47E+06	P53506|ACT8_XENLA	
VAPEEHPVLL	50.61	1102.6022	552.3084	90.5	2.01E+05	P53506|ACT8_XENLA	
VV(sub I)DNGSGMCK	61.91	1008.4368	505.2258	31.08	3.70E+06	P53506|ACT8_XENLA	
YVGDEAE(sub Q)SKRG	60.05	1209.5625	605.7894	26.35	1.16E+05	P53506|ACT8_XENLA	
YVGDEAQSKRG	73.34	1208.5785	605.2971	24.47	1.47E+05	P53506|ACT8_XENLA	
R(+42.01)PQYDASELK	66.01	1247.6146	624.8091	61.6	1.48E+06	tr|D5KU33|D5KU33_BOMMX	Annexin (Bombina maxima)
AQGPP(+31.99)GPAGPAGER	48.2	1292.6108	647.3131	32.56	1.09E+05	tr|O93251|O93251_LITCT	Alpha 1 type I collagen (Lithobates catesbeiana)
GPP(+31.99)GPAGPAGERGEQ	49.93	1407.6378	704.826	30.44	2.64E+05	tr|O93251|O93251_LITCT	
GPP(+31.99)GPAGPAGERGEQGPAG	60.41	1689.7706	845.8885	39.69	1.10E+05	tr|O93251|O93251_LITCT	
AK(+14.02)YGAVPGGAGV	50.49	1059.5713	530.7991	65.24	1.28E+06	tr|Q0Q5Z1|Q0Q5Z1_DANRE	Tropoelastin 1 (Danio rerio)
K(+42.01)PGYGAGAGLIPG	48.2	1198.6345	600.3192	53.6	5.94E+06	tr|Q0Q5Z2|Q0Q5Z2_XENTR	Tropoelastin 1 (Xenopus tropicalis)
EIAQDFKTDL	51.26	1178.5819	590.298	95.32	8.33E+05	tr|Q1WH09|Q1WH09_9NEOB	Histone H3 (Chaperina fusca)

Supplement Table 3

Classification of proteins identified from *Venenum Bufonis* by GO annotation

Class	Protein Name
immune system process	Unigene8539_All
	CL4669.Contig1_All
	Unigene35022_All
	Unigene20657
metabolic process	CL2148.Contig1_All
	CL4920.Contig2_All
	CL4920.Contig3_All
	Unigene20657_All
	Unigene21587
	Unigene30732_All
antioxidant activity	CL127.Contig1_All
	CL127.Contig2_All
	CL127.Contig3_All
	Unigene13919_All