Journal of Venom Research ISSN: 2044-0324 J Venom Res, 2020, Vol 10, 18-22

RESEARCH REPORT

Toxin transcripts in *Crotalus atrox* venom and *in silico* structures of toxins

Ying Jia*, Ivan Lopez and Paulina Kowalski

Biology Department, The University of Texas Rio Grande Valley, Brownsville, Texas 78520, USA

***Correspondence to:** Ying Jia, Email: Ying.jia@utrgv.edu

Received: 16 May 2020 | **Revised:** 15 June 2020 | **Accepted:** 16 June 2020 | **Published:** 17 June 2020

ABSTRACT

The western diamondback rattlesnake (*Crotalus atrox*) is a common and widespread North American pit viper species, and its venom possesses medical applications. In this research, we identified 14 of the most common transcripts encoding 11 major venom toxins including transcripts for a three-finger toxin (3FTx) from the crude venom of *C. atrox*. *In silico* three-dimensional (3D) structures of 9 venom toxins were predicted by using deduced toxin amino acid sequences and a computer programme-MODELLER. The accuracy of all predicted toxin structures was evaluated by five stereochemical structure parameters including discrete optimised protein energy (DOPE) score, root mean square deviation (RMSD), Z-score, overall quality factor (ERRAT), and ϕ/ψ dihedral angle distribution of toxin backbone Cα residues, resulting that the overall predicted models are satisfied quality evaluation checks. Our present toxin transcripts and simulated individual toxin structures are important not only for revealing species-specific venom gene expression profiles, but also for predicting the toxin-toxin interactions and designing the structure-based toxin inhibitors for the treatment of snakebites.

KEYWORDS: *Crotalus atrox*, venom, transcript, toxin, 3D structure

INTRODUCTION

Snake venoms are composed of a diverse array of toxic proteins and peptides (toxins), resulting in a wide variety of pharmacological and toxicological effects (Zhang, 2015). The western diamondback rattlesnake (*Crotalus atrox*) is a common and widespread North American pit viper species. It is also one of the more aggressive rattlesnake species found in North America, and is likely responsible for most snakebite fatalities in northern Mexico and the second greatest number in the U.S. after the eastern diamondback rattlesnake (*Crotalus adamanteus*) (Campbell and Lamar, 2004); therefore, *C. atrox* venom is of considerable clinical importance. Interestingly, although snake venom composition varies interspecifically, intraspecifically and even ontogenetically (Durban, 2013; Gibbs et al, 2013), many abundant venom toxins belong to only a few major toxin families (Gutiérrez et al, 2009). *C. atrox* venom contains many major types of snake venom toxins including snake venom metalloproteinase (SVMP), snake venom serine proteinase (SVSP), phospholipase A_2 (PLA_2), L-amino acid oxidase (LAAO) and C-type lectin-like protein, as well as many lower expressed toxins (Calvete et al, 2009). Therefore, we can use *C. atrox* venom as a representative of snake venoms to unravel the venom complexity. However, to the best of our knowledge, there are only a few incomplete reports with respect to venom gland transcripts of *C. atrox*, but no one so far has systematically analysed the toxin transcripts, particularly from the crude venom of *C. atrox*. In addition, snakebite is a neglected tropical disease that kills or maims hundreds of thousands of people every year, especially in impoverished rural communities of sub-Saharan Africa, Asia, Latin America and parts of Oceania (Gutiérrez et al, 2017). Therefore, due to *C. atrox* venom containing most of the snake venom components, it is vital to fully characterise its venom toxin transcripts and toxic proteins in order to advance our knowledge towards its venom composition in order to pursue an integrated strategy, such as structural based universal anti-venom discovery, to reduce the burden of this neglected tropical disease.

In this report, we cloned 14 of the most common transcripts encoding 11 major toxin families from the crude

venom of *C. atrox* by using one-time reverse transcription polymerase chain reaction (RT-PCR) method. Furthermore, the 3D structures for 9 major toxins were simulated using a template-based computational approach. The acquired 14 venom transcripts are crucial towards predicting venom components and revealing a gene expression profile for single species-specific snake venoms. The simulated 3D toxin structures will play an important role in predicting venom toxin-toxin interactions, as well as in the application of structure-based virtual screening to identify potent inhibitors for the development of universal antivenom. More importantly, the present results demonstrated that venom glands can be substituted by crude venom for characterising at least the abundantly expressed toxin transcripts, which will avoid sacrificing animals for obtaining the venom glands.

MATERIALS AND METHODS

Identification of the most common venom toxin transcripts

To identify the most common toxin transcripts of each toxin family from *C. atrox* crude venom, we adopted the procedure developed in our lab (Jia et al, 2019) with some minor modifications: i) we used pJET1.2 (ThermoFisher Scientific, USA) as the cloning vector; and ii) we divided PCR product (20µl) of each clone into two 10µl, one for running agarose gel to obtain the transcript with correct molecular size and the other 10µl for *Alu*I restriction enzyme digestion of same molecular size to acquire the most common transcript in terms of the enzyme digestion pattern. The recombinant plasmid DNAs isolated from the most common clones were sequenced. Each cDNA sequence was translated into amino acid sequence that was further used for local BLAST against UniProtKB/Swiss-Prot (swissprot) and non-redundant protein sequences (nr) with threshold values identity >30% and E-value $<10^{-5}$ to validate the identity of toxin transcripts. The potential signal peptides were predicted by using Signal4.1 Server.

Homology modelling of venom toxin 3D structures

The deduced amino acid sequence of the most common transcript from each toxin family was used to predict the toxin 3D structures using computer software - MODELLER version 9.23 (Ŝali and Blundell, 1993). Briefly, by performing python script (build_profile.py) (Eswar et al, 2008), the top ranked crystal structures of proteins were retrieved from PDB (Berman et al, 2007) as templates. Based on the alignment confidence (Identity and E-values) and template structure resolution (Å), we selected one crystal structure to carry out an alignment by the script (align2d.py). Further, ten random 3D models were built for each toxin using the model-ligand.py script. The lowest discrete optimised protein energy (DOPE) score model (LDSM) was selected for further refinement. We refined the models using either multiple templates by script (multiple_templae/salign.py) or loop refining by script (loop modelling/loop_refine.py) (Webb and Ŝali, 2016). After refinement, all graphic structures of simulated 3D models were visualised by UCFC ChimeraX programme (Goddard et al, 2018).

All simulated 3D toxin structures were verified by stereochemical parameters: i) The deviation (Root Mean Square Deviation, RMSD) of backbone atoms (Cα) between template protein and predicted toxin was measured by superimposing both structures UCFC ChimeraX; ii) the overall quality factor (ERRAT) was calculated for the model and the template using SAVE v5.0 programme (Laskowski et al, 1993); iii) we utilised the normalised Z-score to determine if the simulated toxin structure falls within range of high-quality experimental structure by using ProSA-web server (Wiederstein and Sippl, 2007); iv) we employed the Ramachandran plot (Ramachandran et al, 1963) to interrogate the ϕ and ψ dihedral angle distributions of toxin backbone Cα residues within the toxin model via PROCHECK (Laskowski et al, 1993).

RESULTS AND DISCUSSION

Identify the most common toxin transcripts

To identify the most common toxin transcripts, we used 100 ng of cDNA reverse-transcribed from venom mRNA as PCR templates, 13 pairs of PCR primers (Table 1) designed for 13 toxin families and High-Fidelity DNA Polymerase to amplify the venom cDNAs, resulting that 13 amplicons (Figure 1) representing 14 different toxin transcripts were obtained. These 14 toxin transcripts include two PLA_2 transcripts (Lys49 and Asp49 PLA_2s) in the same amplicon, three snake venom metalloproteinases (SVMP I, II and III), and one most common transcript for each of the following: LAAO, SVSP, C-type lectin-like protein, three-finger toxin (3FTx), cysteine-rich secretory protein (CRISP), Vespryn, crotamine, epidermal growth factor (EGF) and phospholipase B (PLB). We also tried to amplify transcripts for Nucleotidase and Hyaluronidase but failed probably due to the lower expression of these two transcripts in crude venom. Among these transcripts, PLA_2 is the most highly expressed, followed by LAAO and C-type Lectin-like protein, whereas Crotamine transcript is the least expressed. At the venom toxin translational level, Calvete et al (2009) reported that SVMP and SVSP are the two toxins with the highest expression in the *C. atrox* venom, implying that venom toxin transcripts and toxic proteins in this specific venom underwent unparalleled expression. PLA_2 amplicon contains two bands (upper and lower) (Figure 1); we sequenced both but only the upper band is PLA_2, containing both Lys49 and Asp49PLA_2s, whereas the lower band shows a hypothetic protein transcript. SVMP amplicon consists of three bands, the upper, middle, and lower representing SVMP III, II and I, respectively. Evidently, the SVMP III transcript is the most highly expressed among the three SVMPs. Crotamine amplicon consists of two bands with nearly same molecular size. We sequenced both and they show the same crotamine transcripts. The upper band of vespryn is the most common vespryn transcript, while the lower band shows a rare and truncated vespryn transcript. Although each PCR amplification (lane) contains multiple bands, the anticipated bands for each transcript are predicted in terms of the molecular sizes between two primers.

After each band was purified and individually ligated into pJET vector, and further transformed into *E. coli* competent cells, we obtained numerous clones for each ligation. It is very common that snake venom toxin family contains multiple toxin isoforms encoded by

Table 1. PCR primer sequences for amplifying venom toxin transcripts.

Transcript	Primer pair	Primer sequence (5'-3')
PLA_2	PLA_2 F PLA_2 R	CCGGCTTCTCCTTCTGATCCTT GAGTGCAAAGCTGGCACCTGT
SVMP	SVMP F SVMP R	CCAGCCAAATCCAGCCTCCAAA TGCCCATGGAGCTTTGTG
LAAO	LAAO F LAAO R	TTGAGCACTTTGCTTAGCATCA CTTTCCAAATTGGGGTGGCAT
SVSP	SVSP F SVSP R	GGACACTT CTGGACGTCACT CCCTGCAGCACTATTTTGAGC
C-type lectin	C-type lectin F C-type lectin R	GCCTCTGAGCAGACTTGCTAC ATTTGGACCTTCTGACCCATCTG
3FTx	3FTx F 3FTx R	TGCAGGCTGAAGAGGAGATTG GGGGATTATGGACCATCCTGTT
CRISP	CRISP F CRISP R	CCTGGTACTGTCTGTCTGACTT GCATGAATGGCATCAGATCA
Phospholipase B	PLB F PLB R	ACGGAGGATTCGGCATGAT TGTAAAACACAATTAATCTTCCTGC
EGF	EGF F EGF R	ATGAAAGCCGGTTGGGCTG TCACCTTGGGTAGTGGTTTTTC
Crotamine	Crotamine F Crotamine R	GATGCACCGTTGCCTTAGGT GATACAGCAATAGCAGGCGG
Vespryn	Vespryn F Vespryn R	TCATAGTCTCCAGGGCTCACA AAACGGTCACTAAACCAACGG
Nucleotidase	Nucleotidase F Nucleotidase R	CTCCTTCCTCCGCACTCTTG ATGCCGGCTTCTGAAAGTCT
Hyaluronidase	Hyaluronidase F Hyaluronidase R	CCCTACCTGGTGGATTGGAC GAATTGCCCTCAGATCAAAGC

different transcripts, and these transcripts usually share conserved 5'- and 3'-end untranslated regions but are diverse in open-reading frames (ORFs). To obtain the most common transcript from each toxin family, we screened the clones from each ligation based on the molecular size and enzyme digestion patterns. After screening of at least 46 clones for each ligation, we obtained the most common transcript for each toxin family. Based on enzyme digestion patterns, the most diversified toxin family in *C. atrox* venom is C-type lectin-like protein, whereas the most conserved toxin families are CRISP and PLB. Importantly, to our knowledge, we, for the first time, identified the transcripts encoding a 3FTx from *Crotalus* snake species. The PCR amplicon contains at least two dispersedly duplicated transcripts with a conserved interval (128 nucleotides) (Figure 2), implying that this 3FTx is translated from multiple transcripts. However, further investigation is required to distinguish whether these multiple transcripts were transcribed from the same gene or multiple genes. Based on the length of amino acid sequence and the conserved disulphide bond patterns (Kini and Doley, 2010, Kessler et al, 2017), this 3FTx belongs to the short neurotoxins (64 amino acids) targeting various receptor/ion channel proteins such as nicotinic acetylcholine receptors (Kini, 2011; Kessler et al, 2017), and it shows 86% identity with a 3FTx (ABZ89717) from *Sistrurus catenatus edwardsi* (Doley et al, 2008).

The deduced amino acid sequences of all the most common transcripts were blasted against Swiss-Prot and then non-redundant databases to search the homologues in *Crotalus* species, resulting that the identity between query and subject toxins ranged from 44 to 100%, with the highest (100%) of such as PLA_2 and the lowest (44%) of crotamine, except no reports to date for 3FTx from *Crotalus* species (Supplementary Table 1). All deduced toxin amino acid sequences were deposited in NCBI GenBank with accession numbers (MN506247-MN506258) except two PLA_2s, which are identical with previously submitted PLA_2 sequences (APD70896 and Q8UVZ7).

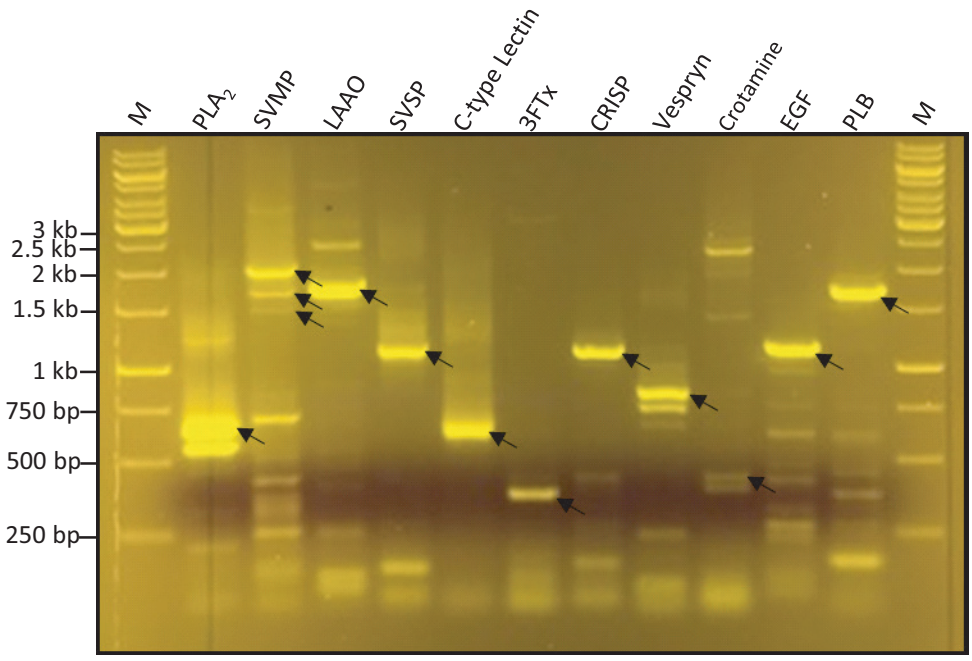

Figure 1. RT-PCR. Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, USA) and 100 ng of cDNA revise-transcribed from venom mRNA were used to amplify each venom toxin transcript with primers listed in Table 1 at the annealing temperature of 55 °C for 40 cycles. The expected amplicons arrowed were predicted based on the molecular size between two primers, purified and ligated into pJET1.2 vector (ThermoFisher Scientific, USA) for identifying the most common transcript of each venom toxin family.

***In silico* 3D structure of toxin**

The availability of a venom toxin's structural model is one of the keys to understanding the toxic activity at the molecular level. We individually blasted all deduced amino acid sequences of 14 toxin transcripts against Protein Data Bank (PDB) (Berman et al, 2007), and found 9 experimentally obtained 3D structures of protein with the lowest identity of 41.8% (3FTx) to the highest of 99.8% (SVMP III). Supplementary Table 2 shows the 9 simulated venom toxin structures and superimposition of toxin model (in different colours) on the protein template coloured in magenta, as well as the structure validation. A total of 10 models were generated for each toxin by MODELLER, and the model with the lowest DOPE score was selected for quality inspection. The inspection of DOPE score energy profiles indicated that the regions of some venom toxins such as C-type lectin-like protein and 3FTx showed higher energy. Thus, the predicted structures of such toxins were further refined by using either multiple templates or loop refinement methods. After refinement, the predicted structures were vastly improved when compared with template structures.

The estimation of the accuracy and reliability of simulated structures was one of the most important steps in protein structure simulation. Therefore, we employed various stereochemical parameters including RMSD, ERRAT, Z-Score, and backbone ϕ/ψ angles to validate the predicted toxin structures: i) the quality of modelled structures was evaluated by computing the RMSD of backbone atoms of the model vs the template. The RMSD ranged from 0.18 Å to 0.58 Å, with the highest of PIII and 3FTx (0.58 Å) and lowest of Asp49 (0.18 Å), which are significantly lower than the threshold value of 2 Å; ii) ERRAT is a so-called "overall quality factor" (Colovos and Yeates, 1993) for non-bonded atomic interactions, with higher scores indicating higher quality. ERRAT scores for all predicted toxins were calculated to be from the lowest of 76.7 for PLB to the highest of 92.1 for Lys49, which is above the expected accuracy threshold of 70% for medium resolution structures; iii) the energy capacity to be accurate for simulated toxins was calculated in terms of Prosa Z-score (Wiederstein and Sippl, 2007); all Z scores for models are negative and were arranged for the least comparable for 3FTx and template (2VLW) from -3.4 to -4.3, and the most comparable for PLB and template

Figure 2. Dispersed multiple transcripts of three-finger toxin (3FTx). The translated cDNA sequence by ExPASy server shows that the readable sequence of one Sanger sequencing reaction contains two 3FTx open reading frames and a short 3FTx fragment (dark background) adjacent by 128-nucleotide intervals of conserved untranslated regions.

(4BWC) equal to -6.7, suggesting that the models fall in the range of at least medium quality crystal structures of a similar size and shape; iv) Ramachandran plot (ϕ, ψ-plot) analysis indicated that more than 85% of residues for predicted toxin structures except 3FTx (78.2%) were found in most favoured regions. More than 6% of residues were found in additional allowed regions, meaning that more than 90% of residues in the toxin structures satisfied phi and psi dihedral angle distributions. The predicted 3FTx structure with the 78.2% residues falling in favoured region is probably due to either the template (2VLW) with 87.5% residues in favoured region, or the lower identity (41.8%) between the template (2VLW) and 3FTx. Overall, the results shown in Supplementary Table 2 support that all predicted toxin models are of good quality, and satisfied the quality checks.

CONCLUSIONS

Elucidating the mechanism of venom toxin function generally requires knowledge of toxin structure, which is determined by toxin amino acid sequence. Currently, a large number of venom toxin transcript, and even native toxin, sequences are available. However, there is a large gap between the number of available toxin sequences and their experimentally solved toxin structures. To fill in this gap, the homology modelling can play a central role. In the present work, we identified 14 of the most common toxin transcripts including, for the first time, a 3FTx transcript from the *Crotalus* snake species. These transcripts are important for displaying toxin gene expression profiles, dissecting toxin gene structures, and predicting venom toxin components. We believe that the simulated 3D structures of 9 major toxins will be important in developing structure-based toxin inhibitors for the treatment of snakebites, predicting toxin-toxin interactions to uncover snake venom complexity, as well as conducting similar studies for other animal venoms, such as spider and scorpion venoms.

ACKNOWLEDGMENTS

Authors are grateful for support from the UTRGV High Scholars Programme and Engaged Scholar Award at the University of Texas Rio Grande Valley.

COMPETING INTEREST

None declared.

ABBREVIATIONS

DOPE: discrete optimised protein energy
RMSD: root mean square deviation
ERRAT: overall quality factor
SVMP: snake venom metalloproteinase
SVSP: snake venom serine proteinase
PLA_2: phospholipase A_2
3FTx: three-finger toxin
LAAO: L-amino acid oxidase
CRISP: cysteine-rich secretory protein
EGF-like: epidermal growth factor-like domain protein
PLB: phospholipase B
RT-PCR: Reverse transcription polymerase chain reaction
ORF: open-reading frames
PDB: Protein Data Bank

REFERENCES

Berman H, Henrick K, Nakamura H and Markley JL. 2007. The worldwide protein data bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res, 35, D301-D303.

Calvete JJ, Fasoli E, Sanz L, Boschetti E and Righetti PG. 2009. Exploring the Venom Proteome of the Western Diamondback Rattlesnake, *Crotalus atrox*, via Snake Venomics and Combinatorial Peptide Ligand Library Approaches. J Proteome Res. 8, 3055-3067.

Campbell JA and Lamar WW. 2004. The venomous reptiles of the Western Hemisphere: Comstock Publishing Associates: Ithaca, NY, 2004.

Colovos C and Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci, 2, 1511-1519.

Doley R, Pahari S, Mackessy SP and Kini RM. 2008. Accelerated exchange of exon segments in Viperid three-finger toxin genes (*Sistrurus catenatus edwardsii*; Desert Massasauga), BMC Evolutionary Biol, 8, 196.

Durban J. 2013. Intergrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake. *Crotalus simus simus*. BMC Genomics, 14, 234.

Eswar N, Eramian D, Webb B, Shen MY, Shen MY and Sali A. 2008. Protein structure modelling with MODELLER. Methods Mol Biol, 426.

Gibbs HL, Sanz L, Sovic MG and Calvete JJ. 2013. Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (*Sistrurus* sp.). PLoS ONE, 8, e67220.

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci, 27, 14-25.

Gutiérrez JM, Lomonte B, Leon G, Alape-Giron A, Flores-Diaz M, Sanz L, et al. 2009. Snake venomics and antivenomics: proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming. J Proteomics, 72, 165-182.

Gutiérrez JM, Calvete JJ, Habib AG, Harrison RA, Williams DJ and Warrell DA. 2017. Snakebite envenoming. Nat Rev Dis Primers, 3, 17079.

Jia Y, Olvera P, Rangel F, Mendez B and Reddy S. 2019. Rapid identification of phospholipase A_2 transcripts from snake venoms. Toxins, 11, 69.

Kessler P, Marchot P, Silva M and Servent D. 2017. The three-finger toxin fold: a multifunctional structural scaffold able to modulate cholinergic functions. J Neurochem, 142, 7-18.

Kini RM and Doley R. 2010. Structure, function and evolution of three-finger toxins: Mini proteins with multiple targets. Toxicon, 56, 855-867.

Kini RM. 2011. Evolution of three-finger toxins – a versatile mini protein scaffold. Acta Chim. Slov. 58, 693-701.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structure. J Appl Cryst, 26, 283-291.

Ramachandran GN, Ramakrishnan C and Sasisekharan V. 1963. Stereochemisty of polypeptide chain configurations. J Mol Biol, 7, 95-99.

Ŝali A and Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol, 234, 779-815.

Wiederstein M and Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res, 35, W407-W410.

Webb B and Ŝali A. 2016. Comparative protein structure modelling using MODELLER. Curr Protoc Bioinformatics, 54, 5.6.1-5.6.37.

Zhang Y. 2015. Why do we study animal toxins? Dongwuxue Yanjiu, 36, 183-222.