Austine James et al /J. Pharm. Sci. & Res. Vol. 11(7), 2019, 2543-2548

ISSN:0975-1459
Journal of Pharmaceutical Sciences and Research
www.jpsr.pharmainfo.in

A Computational Strategy to Predict Lethal Activity of Snake Venom Cardiotoxins

Austine James, K. Sudharshan, S. Kishore and T. Sivaraman*

Drug Design and Discovery Laboratory, Department of Biotechnology,
Karpagam Academy of Higher Education, Coimbatore – 641021, Tamil Nadu, India.

Abstract

Venoms from *elapidae* snakes are rich sources of over 100 protein toxins and of which, cardiotoxins (CTXs) are most abundant protein toxins. To date, 88 authentically annotated primary structures of the CTXs and 20 experimental three-dimensional (3D) structures of the CTXs from various species of snakes have been reported in the literature. Notwithstanding their sequence and structural similarities, lethal doses (LD_{50}) of the CTXs have been found to be varied from 0.8 to 4.9 mg/Kg b.w. (i.v). The LD_{50} values of the CTXs are being determined using authentic functional assays, which prerequisites highly pure protein samples and sound experimental knowledge and the correlations between the LD_{50} values and CTXs would help to annotating these protein toxins. In this context, we have developed a computational approach to predict LD_{50} of the CTXs on the basis of unique BLOCKS computed from their primary structures. The robustness of the computational approach has been validated using authentic sequences of the CTXs for which LD_{50} values are reported in the literature and moreover, the merits and scopes of the approaches have also been discussed in a concise manner.

Keywords: Cardiotoxins, Hemolysis, Lethal dose, Three-finger folds and Snake venom.

INTRODUCTION

Around 3400 snake species have been reported to date and about 600 species of them are identified as venomous snakes [1,2]. The snake venom consists of tens of peptides, proteins and enzymes and the protein toxins have been grouped into six superfamilies [3]. Of the six superfamilies, three-finger toxin (3FTx) superfamily contains most toxic principles and cardiotoxins (CTXs) are most abundant protein toxins of the 3FTx [3-6]. To date, 88 authentically annotated primary structures of the CTXs purified from various species of snakes have been reported in the literature. As on April 2019, 20 three-dimensional (3D) structures of the CTXs determined by experimental methods have been deposited in the 'protein data bank' and the CTXs depict simple β-sheet folds: five anti-parallel strands, three loops, a globular head and an unstructured C-terminal segment [3,5-8].

All the CTXs of snake venoms reported to date are very similar to one another in terms of their primary, secondary and tertiary structures [1,7,9]. Notwithstanding their structural similarities, they are drastically differing from one another in their LD_{50} activities (lethal dose). Authentic functional assays are prerequisites for unambiguously annotating these protein toxins and the assays demand highly pure protein samples and sound experimental knowledge [10-13]. In this background, we have developed a computational approach to predict LD_{50} of the cardiotoxins on the basis of unique BLOCKS computed from their primary structures [14,15]. The robustness of the computational approach has been validated using authentic sequences of the cardiotoxins for which LD_{50} values are reported in the literature. Moreover the merits and scopes of the approaches have also been discussed in a concise manner.

METHODS

Sequence collections, annotations and alignments

Amino acid sequences of snake venom CTXs were retrieved from NCBI protein database (https://www.ncbi.nlm.nih.gov/) and as well from UniProt database (https://www.uniprot.org/). The CTXs were confirmed on the basis of their sequence annotations and as well by using TFTX tool (http://sblab.sastra.edu/tftx.html). The TFTX is a computational tool and predicts whether the given sequence is snake venom CTX or not and robustness of the tool has been documented in the literature [4]. MultAlign tool (http://multalin.toulouse.inra.fr/multalin/) was used to perform multiple sequence alignments for the 88 non-redundant CTX sequences considered in the present study and the alignment parameters were set as follows: Matrix: BLOSUM62; Gap Weight: 12; Gap Length: 2; Consensus levels: 50% [16]. The 21 CTXs for which LD_{50} values were available in the literature were grouped into three categories I, II, & III based on their LD_{50} values ranging from 0.8 to < 2.0, 2.0 to < 3.0, 3.0 to 4.9, respectively. The same sequence alignment parameters mentioned above were used to identify unique BLOCKS (a short segment) of the CTXs belonging to each of the three categories.

Phylogenetic analysis

Phylogenetic tree for the 88 CTXs considered in the present study was constructed using MEGA 7 and the statistical method used for the purpose was Maximum likelihood method [17-19]. The generated tree was subjected to further refinements using 'Bootstraping' method: the number of iterations and substitution matrix were 500 and Dayhoff model, respectively. All the CTXs are referred by their corresponding accession numbers in the phylogenetic tree. The LD_{50} values available for the 21

CTXs have been indicted within parenthesis just next to their corresponding accession IDs or accession numbers.

RESULTS AND DISCUSSION

Snake venom cardiotoxins (CTXs) are single polypeptide chain consisting of 59 – 62 standard amino acids and they are basic proteins (pI > 9.5). The CTXs belonging to three-finger toxin superfamily are abundant and most toxic principal component of elapid snake venoms [7]. As illustrated in the multiple sequence alignments (Figure 1), all CTXs, except two sequences, begin with 'leucine' and end with a pair of 'Cysteine – Asparagine' residues. Also, all CTXs have conserved 'cysteine' and 'proline' residues (Cys3, Pro8, Cys14, Gly17, Leu20, Cys21, Arg36, Cys38, Pro43, Ser46, Cys53, Cys54 and Cys59) and as well signature peptides [20]. As far as considering secondary structures, all CTXs are β-sheet proteins and around 50% of the residues of the proteins are responsible for constituting their secondary structural frameworks. Overall tertiary structures of the CTXs are very similar, if not identical, to one another: three loops (Loop I, II & III) are protruded from the globular head of the proteins and hence the name 'three-finger protein toxins'; the 'Loop I' is constituted by stand I & II (double stranded domain); Loop II & III are constituted by stand III, IV & V (triple stranded domain). Taken together, all CTXs are highly similar to one another in terms of their primary, secondary and tertiary structural contexts.

Irrespective of similar structural architectures, the CTXs depict differential binding modes with lipid bilayer and accordingly the CTXs have been classified into three types: P-type CTXs, SL-type CTXs and SK-type CTXs [9,21,22]. Interestingly, the CTXs are also differing from one another in their LD_{50} (lethal dose through intravenous mode) values estimated from the animal model experiments for the protein toxins as reported in the literature [23-29]. As on April 2019, LD_{50} values for 21 CTXs have been reported by various research groups worldwide and the values have been ranged from 0.83 mg/Kg b.w. to 4.9 mg/Kg b.w. (about six-fold difference). In order to understand the evolutionary relationships for the variations in the LD_{50} values of the CTXs, a phylogenetic tree was generated for all the 88 CTXs available to date as described in the method section (Figure 2). From a quick inspection to the phylogenetic tree of the CTXs, it is obvious that there were no clear connections between the amino acid sequences and the LD_{50} values of the snake venom CTXs.

Table 1: Classification of 21 CTXs (for which experimental LD_{50} values are available) into three groups on the basis of 'Loop II BLOCKS' of the protein toxins. All CTXs are represented by their accession numbers and the accession numbers of the CTXs that deviate from the prediction are shown in red colour. The LD_{50} values of the CTXs are given in parenthesis next to their corresponding accession numbers.

Group I 0.8 ± 0.1 to < 2.0 ± 0.2 mg/Kg b.w. (i.v)	Group II 2.0 ± 0.2 to < 3.0 ± 0.3 mg/Kg b.w. (i.v)	Group III 3.0 ± 0.3 to 4.9 ± 0.5 mg/Kg b.w. (i.v)
P01467 (0.8); P01469 (1.1); P01445 (1.2); P01446 (1.2); P60305 (1.3); P01448 (1.4); P01470 (1.8); P01452 (2.0); P01442 (2.1); P01443 (2.1); P25517 (2.9); P24776 (3.8).	P01464 (1.8); P01462 (2.0); P24777 (2.5).	P01460 (2.6); P01459 (3.0); P01455 (3.0); P01461 (4.0); P01453 (4.1); P01454 (4.9).

Table 2: Classification of 67 CTXs (for which experimental LD_{50} values were not available) into three groups on the basis of 'Loop II BLOCKS' of the protein toxins. All the CTXs are represented by their accession numbers.

Group I 0.8 ± 0.1 to < 2.0 ± 0.2 mg/Kg b.w. (i.v)	Group II 2.0 ± 0.2 to < 3.0 ± 0.3 mg/Kg b.w. (i.v)	Group III 3.0 ± 0.3 to 4.9 ± 0.5 mg/Kg b.w. (i.v)
O73856; O93473; P24779; P60309; P01441; Q9PS34; O73857; P60310; Q98961; O73858; O73859; P60311; P07525; Q9W6W9; P0CH80; P83345; P80245; P49123; Q98965; Q9W6W6; P01468; P01457; P62375; Q91996; Q9W716; Q8UUK0; Q91126; P49122; P01471; B3EWH9 P60306; P24780; P01447; P86540; P86382; P60304; Q98958; Q91135; Q98957; P79810; Q91136; P01451; P86538; Q98956	P60301; Q02454; P60302; P60303; Q98959; Q9PST4; P01466; Q9PST3; O93471; P86541; P01440; Q98960; P01465; O93472; Q91124; Q98962; P01463; Q9DGH9; P60307; P60308 A0A0U5AUY6	P01468; P01458

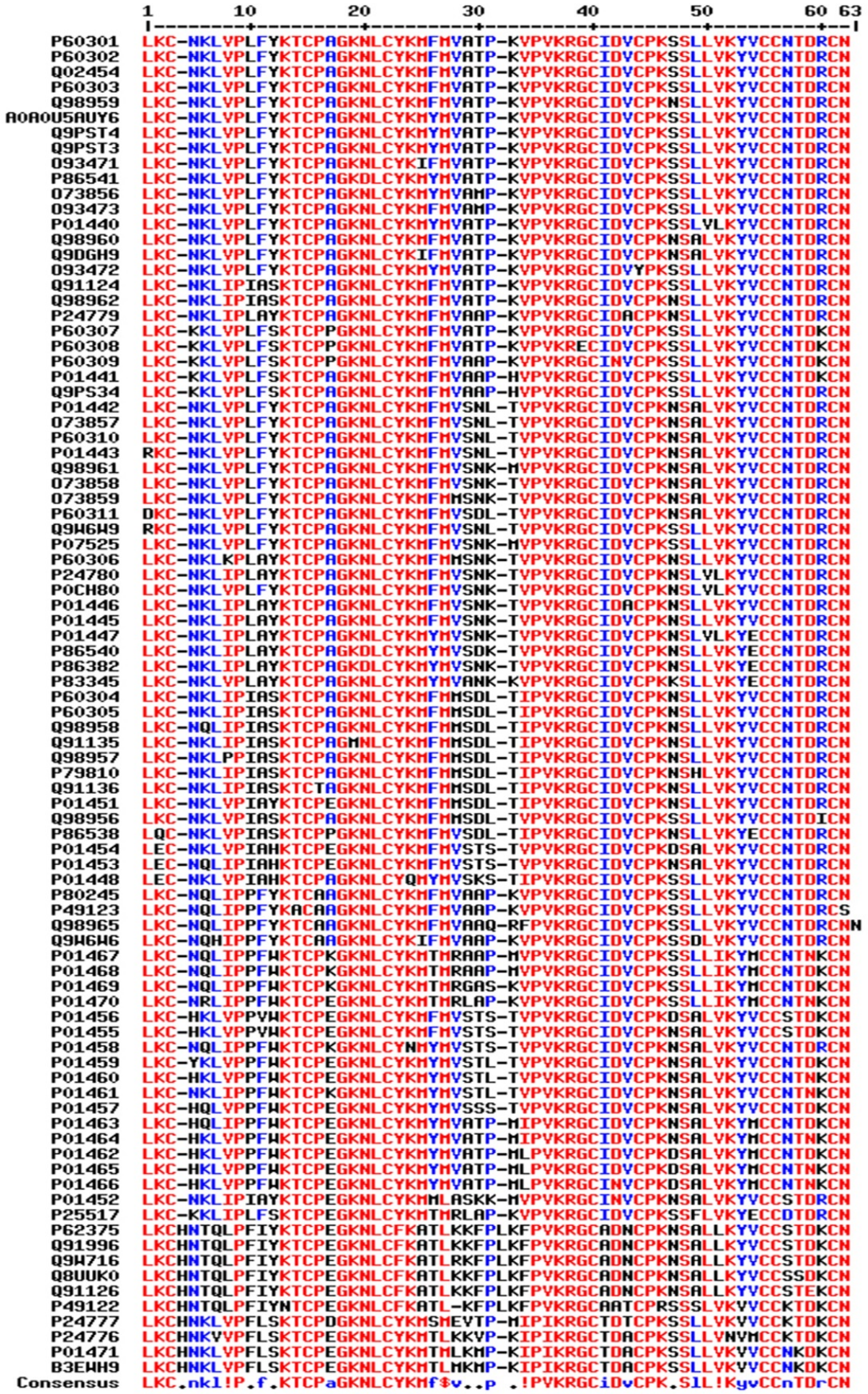

Figure 1: A figurative representation showing multiple sequence alignments of the 88 cardiotoxins from various snake venoms as obtained from MultAlign computational tool.

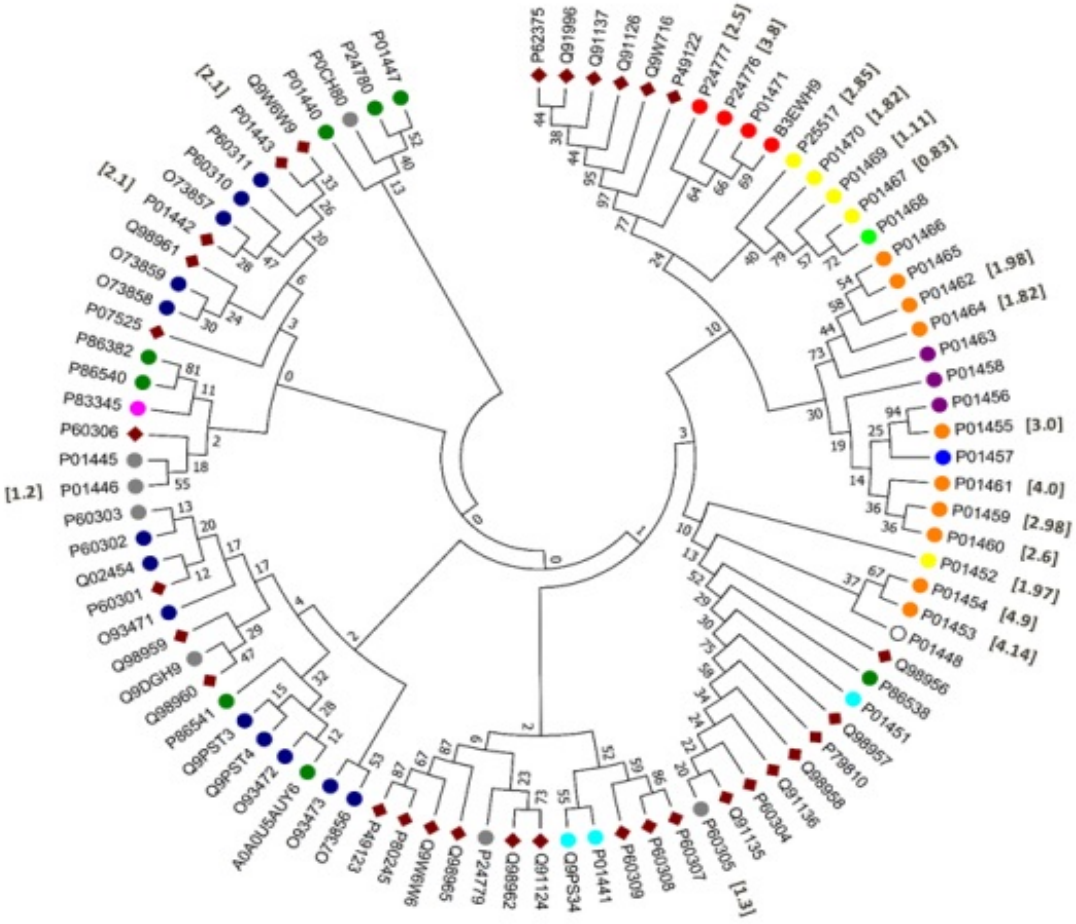

Figure 2: Phylogenetic tree generated using MEGA 7 computational tool for the 88 non-redundant and authentic snake venom cardiotoxins deposited in the primary database is depicted. LD_{50} values for 21 of the 88 cardiotoxins have been denoted in parenthesis beside to their corresponding accession numbers.

As mentioned above, overall secondary and tertiary structures of all the CTXs are very similar to one another. In this background, in order to find relationships between the sequences - LD_{50} values of the CTXs, we attempted to identify unique 'BLOCKS' in the primary structures of the CTXs. For sake of understanding the particular relationships, the 21 CTXs for which LD_{50} values are available in the literature were divided into three groups: the LD_{50} values for CTXs belonging to the Group I were ranged from 0.83 mg/Kg b.w. to < 2.0 mg/Kg b.w.; the LD_{50} values for the CTXs belonging to the Group II were ranged from 2.0 mg/Kg b.w. to < 3.0 mg/Kg b.w.; the LD_{50} values for CTXs belonging to the Group III were ranged from 3.0 mg/Kg b.w. to 4.9 mg/Kg b.w. In the meantime, on the basis of error associated with the LD_{50} values reported for a few CTXs, the upper and lower limits set for the three Groups were allowed with 10% liberalization. The accession IDs of the CTXs belonging to the three groups are as mentioned herein: Group I - P01467, P01469, P01445, P01446, P60305, P01448, P01470, P01464, P01452 and P01462; Group II - P01442, P01443, P24777, P01460, P25517 and P01459; Group III - P01455, P24776, P01461, P01453, P01454. On the basis of the sequence alignments, two BLOCKS were identified for each of the three groups. One BLOCKS was constituted at positions 10 & 11 and another BLOCKS was constituted at positions 30, 31 & 32. From three-dimensional structural standpoints, the former and latter BLOCKS were situated in the tips of Loop I and Loop II of the CTXs, respectively. The Loop I BLOCKS for the Group I, II and III CTXs were AX, FS and XX, respectively. Similarly, the Loop II BLOCKS for the Group I, II and III CTXs were XXX, TPX and TXT, respectively. In both the cases, the letter 'X' stands for anyone of the twenty standard amino acids. It should also be mentioned that the position mentioned for the residues constituting the Loop I and II BLOCKS would vary for the CTXs possessing 'Histidine' at 4th position and as well CTXs possessing 'Leucine' at 32nd position (Figure 1).

The CTXs for which LD_{50} values are available were classified on the basis of the two 'BLOCKS' identified in the present study. The classifications of the 21 CTXs carried out either on the basis of 'Loop I BLOCKS' alone or combination of 'Loop I & II BLOCKS' were not reliable (data not shown). On the other hand, the classifications carried out for the CTXs on the basis of the

'Loop II BLOCKS' were reliable with 86% confidence limits (Table 1). In this context, it is also interesting to point out that primary functional unit of the CTXs is presumably Loop II to elicit their biological activities such as cytolysis, haemolysis, cardiac muscle damages and membrane depolarization. Of the 21 sequences, 18 sequences could be correctly classified into their corresponding groups; the three CTXs (P25517, P24776 and P01460) were found to deviate from the classification by 'Loop II BLOCKS'. The three sequences could not be reliably classified using the former strategy too ('Loop I/II BLOCKS') implying that the three sequences presumably demands further experimental and computational validations in an accurate manner for authentic classifications.

In this background, the 67 CTXs for which LD_{50} values have not yet been deposited in the public domain to date were subjected to classifications in order to predict their LD_{50} values on the basis of 'Loop II BLOCKS' and the results are shown in Table 2. The 67 sequences were classified into 44, 21 and 2 into Group I, Group II and Group III, respectively. The data suggested that the LD_{50} values for the 44 sequences belonging to Group I have been predicted to be ranged from 0.8 ± 0.1 to 2.0 ± 0.2 mg/Kg b.w. (i.v); similarly the LD_{50} values for the 21 sequences belonging to Group II would be varied from 2.0 ± 0.2 to 3.0 ± 0.3 mg/Kg b.w. (i.v) and for 2 sequences belonging to Group III would be varied from 3.0 ± 0.3 to 4.9 ± 0.5 to mg/Kg b.w. (i.v) as predicted by the strategies described in the present study. In this context, it would be preferable to validate/confirm the predicted LD_{50} values for the 67 CTXs authentically deposited in the primary databases (NCBI/UniProt) through experimental methods.

CONCLUDING REMARKS

In the present study, we have developed a computational strategy to predict LD_{50} values of snake venom CTXs in a systematic manner. The computational strategies have been validated using the 21 CTXs for which LD_{50} values have been authentically reported in the literature and the comprehensive analyses suggested that robustness of the strategy is about 86%. Using the computational approach, LD_{50} values for the non-redundant 67 CTXs (for which LD_{50} values were not available in the literature) have been predicted and rationalized on the basis of unique BLOCKS identified in the primary structures of the CTXs.

Acknowledgement:

The authors would like to express their sincere thanks to the management of Karpagam Academy of Higher Education for providing computational facilities to successfully complete the *in silico* experiments of the present study.

REFERENCES

[1] Hegde, R.P., Rajagopalan, N., Doley, R., Kini, R.M. (2009). Snake venom three-finger toxins. In: Mackessy SP (ed) Handbook of venoms and toxins of reptiles. CRC, Boca Raton, pp 287–302

[2] Kini, R.M. (2006). Anticoagulant proteins from snake venoms: structure, function and mechanism. BiochemJ 397:377–387. doi:10.1042/BJ20060302

[3] Kini, R.M., Doley, R. (2010). Structure, function and evolution of three-finger toxins: mini proteins with multiple targets. Toxicon 56:855–867. doi:10.1016/j.toxicon.2010.07.010

[4] Rajesh, S.S., and Sivaraman, T. (2011). TFTX: A computational tool for predicting subfamilies of three-finger toxins from the venoms of elapid snakes, J. Pharm. Sci. and Res. 3(12), 1612-1618.

[5] Tsetlin, V. (1999). Snake venom alpha-neurotoxins and other 'three-finger' proteins. Eur. J. Biochem. 264, 281–286.

[6] Menez, A. (1998). Functional architectures of animal toxins: a clue to drug design? Toxicon 36, 1557–1572.

[7] Kumar, T.K.S., Jayaraman, G., Lee, C.S., Arunkumar, A.I, Sivaraman, T., Samuel, D., & Yu, C. (1997). Snake Venom Cardiotoxins-Structure, Dynamics, Function and Folding. Journal of Biomolecular Structure and Dynamics, 15(3), 431–463. doi:10.1080/07391102.1997.10508957

[8] Gorai, B., & Sivaraman, T. (2013). Unfolding stabilities of two paralogous proteins from *Naja naja naja* (Indian cobra) as probed by molecular dynamics simulations. Toxicon, 72, 11–22.doi:10.1016/j.toxicon.2013.05.024

[9] Gorai, B., Karthikeyan, M., and Sivaraman, T. (2016). Putative membrane lytic sites of P-type and S-type cardiotoxins from snake venoms as probed by all-atom molecular dynamics simulations. Journal of Molecular Modeling, 22(10). doi:10.1007/s00894-016-3113-y

[10] Sivaraman, T., Kumar, T.K.S., Hung, K.W., and Yu, C. (1999). Influence of disulfide bonds on the induction of helical conformation in proteins. J. Prot. Chem. 18, 481-488.

[11] Jayachandra, K., Sivaraman, T. (2011). Hepatoprotective Effect of *Aegle Marmelos (L.) Corr*. Leaf Powder (Crude) Against Carbon Tetrachloride-Induced Hepatic Damage in Albino Rats. *Journal of Pharmaceutical Sciences and Research*, 3(7), 1360-1363.

[12] Oukkache, N., Jaoudi, R.E. (2014). Evaluation of the lethal potency of scorpion and snake venoms and comparison between intraperitoneal and intravenous injection routes. Toxins 6: 1873-1881.

[13] Tohamy, A.A., Mohamed, A.F. (2014). Biological effects of *Naja haje* crude venom on the hepatic and renal tissues of mice. J King Saud Uni-Sci26: 205-212.

[14] David, W. Mount., Bioinformatics: Sequence and Genome Analysis. 2004. 8th ed., CSHL Press.

[15] Attwood, Teresa. K., David Parry-Smith. Introduction to Bioinformatics. 1999. Addison Wesley Longman limited.

[16] Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering, Nucl. Acids Res., 16 (22), 10881-10890

[17] Schwarz, R., and Dayhoff, M. (1979). Matrices for detecting distant relationships. In Dayhoff M., editor, Atlas of protein sequences, pages 353-58. National Biomedical Research Foundation.

[18] Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874.

[19] Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

[20] Gayatri, S., Gorai, B., Sivaraman, T. (2016). Understanding the Structures and Functions of Snake Venom Cardiotoxins. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 7(2), 936 - 9 41.

[21] Dubovskii, P.V., Lesovoy, D.M., Dubinnyi. M.A., Konshina, A.G., Utkin, Y.N., Efremov, R.G., Arseniev, A.S. (2005). Interaction of three-finger toxins with phospholipid membranes: comparison of S- and P-type cytotoxins. Biochem J 387:807–815. doi:10.1042/bj20041814

[22] Suzuki-Matsubara, M., Athauda, S.B.P., Suzuki, Y., Matsubara, K., & Moriyama, A. (2016). Comparison of the primary structures, cytotoxicities, and affinities to phospholipids of five kinds of cytotoxins from the venom of Indian cobra, *Naja naja*. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 179, 158–164. doi:10.1016/j.cbpc.2015.09.015

[23] https://www.uniprot.org/

[24] Joubert, F. J., & Taljaard, N. (1980). The complete primary structures of three cytotoxins (CM-6, CM-7 and CM-7A) from *Naja naja kaouthia* (Siamese cobra) snake venom. Toxicon, 18(4), 455–467. doi:10.1016/0041-0101(80)90053-7

[25] Das, T., Bhattacharya, S., Halder, B., Biswas, A., Das Gupta, S., Gomes, A., & Gomes, A. (2011). Cytotoxic and antioxidant property of a purified fraction (NN-32) of Indian *Naja naja* venom

on Ehrlich ascites carcinoma in BALB/c mice. Toxicon, 57(7-8), 1065–1072. doi:10.1016/j.toxicon.2011.04.012

[26] Joubert, F.J. (1977). Snake Venom Toxins. The Amino-Acid Sequences of Three Toxins (9B, 11 and 12A) from *Hemachatus haemachatus* (Ringhals) Venom. European Journal of Biochemistry, 74(2), 387–396. doi:10.1111/j.1432-1033.1977.tb11403.x

[27] Joubert, F.J. (1976). Snake venom toxins. The amino acid sequence of three toxins (CM-2e, CM-4a and CM-) from *Naja haje annulifera* (Egyptian cobra) venom. Hoppe-Seyler's Z. Physiol. Chem. 357:1735-1750

[28] Joubert, F.J. (1976). Snake Venom Toxins. The Amino-Acid Sequences of Three Toxins (CM-8, CM-11 and CM-13a) from *Naja haje annulifera* (Egyptian cobra) Venom. Eur. J. Biochem. 64:219-232

[29] Kaneda, N., Sasaki, T., & Hayashi, K. (1977). Primary structures of cardiotoxin analogues II and IV from the venom of *Naja naja atra*. Biochimica et Biophysica Acta (BBA) - Protein Structure, 491(1), 53–66. doi:10.1016/0005-2795(77)90040-x