BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access

www.bioinformation.net

Hypothesis

Volume 8(18)

Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates

Subhamay Panda[1] & Goutam Chandra[2]*

[1]Division of Biological Sciences, Department of Pharmacy, Gupta College of Technological Sciences, Ashram More, G.T. Road, Asansol-713301, West Bengal, India; [2]Department of Zoology, The University of Burdwan, Golapbag, Burdwan -713104, West Bengal, India; Goutam Chandra D. Sc – Email: goutamchandra63@yahoo.co.in; *Corresponding author

Received August 31, 2012; Accepted September 03, 2012; Published September 21, 2012

Abstract:
Snake venom contains a diverse array of proteins and polypeptides. Cytotoxins and short neurotoxins are non-enzymatic polypeptide components of snake venom. The three-dimensional structure of cytotoxin and short neurotoxin resembles a three finger appearance of three-finger protein super family. Different family members of three-finger protein super family are employed in diverse biological functions. In this work we analyzed the cytotoxin, short neurotoxin and related non-toxin proteins of other chordates in terms of functional analysis, amino acid compositional (%) profile, number of amino acids, molecular weight, theoretical isoelectric point (pI), number of positively charged and negatively charged amino acid residues, instability index and grand average of hydropathy with the help of different bioinformatical tools. Among all interesting results, profile of amino acid composition (%) depicts that all sequences contain a conserved cysteine amount but differential amount of different amino acid residues which have a family specific pattern. Involvement in different biological functions is one of the driving forces which contribute the vivid amino acid composition profile of these proteins. Different biological system dependent adaptation gives the birth of enriched bio-molecules. Understanding of physicochemical properties of these proteins will help to generate medicinally important therapeutic molecules for betterment of human lives.

Keywords: Snake venom, Cytotoxin, Short neurotoxin, Three-finger proteins, Bioinformatics, Physicochemical characterization

Background:
Widely accepted view related to phylogeny of snakes that they evolved during the era of dinosaurs in the Jurassic period from a family of terrestrial lizards about 200 million years (Myr) ago **[1, 2]**. Venom, the advanced thesaurus of secretion of venom gland is usually used by snakes in defense and in assault. Within the natural world the venom system of snakes is an example of ultimate sophistication of integrated armory **[3].** Natural selection gives the birth of venom which is a naturally engineered lethal admixture of peptides and proteins. The venom helps snakes to affect different prey or victim by exerting action upon different vital system **[1].** Snake-bites are one of the serious public health problems in many countries of the world. At global level there are 5 million snake-bites, 2.5 million envenoming and over 125,000 mortality annually **[4].** In India the incidence of snake-bites is nearly 200,000 and 35,000-50,000 people are died every year [**5**]. The deadly venom contains plethora of polypeptide and non-polypeptide constituents. Cytotoxins and short neurotoxins are the non-enzymatic polypeptides (Molecular weight 5-10 kDA) within in the snake venom **[6].** Interestingly the cytotoxins and short neurotoxins are the family members of 'Three-finger' protein

(TFP) superfamily. The naming of 'Three-finger' is for its appearance of three loops (finger like) projected from the core region of the protein. Three finger appearances are maintained by three disulfide bridges within the loops **[7].** Cytotoxin exerts their effect upon the target cells by formation of pore within the cell membrane **[8].** Short neurotoxins block the neuromuscular transmission by selective binding to muscle nAChR **[9]**. Other non-toxin family members of Three-finger protein (TFP) superfamily are xenoxin, CD59, Ly-6, Lynx-1 **[3, 7]**. Xenoxin is a skin secretory protein of *Xenopus laevis* frog, CD59 is a complement regulatory protein plays a role in complement system in human, mouse and rat **[7].** Lynx-1 is a neuronal modulator acts on CNS in mouse **[10].** Venom proteins of snakes evolve from the genes of normal body proteins which are responsible for key regulatory processes within the body. These genes are duplicated and selective expression of these duplicated genes facilitates the synthesis of venomous composition of venom gland. In this process the ancestral function is converted into a derived one **[3].** The objective of the present study is a comparative compositional, physicochemical characteristics and functional analysis of snake venom toxin proteins and non-toxin proteins of other chordates like hagfish, frog, mouse, rat and human etc. These comparative analyses will help us to understand the occurrence of diversification of different protein sequences in these toxin proteins and non-toxin proteins of other chordates. This also hints the system-level adaptability of these three-finger proteins in different physiological milieu. From the applicability view point, the results will provide information necessary for generation of engineered therapeutic proteins from the natural toxins.

Methodology:

Amino acid sequences of proteins were obtained from National Centre for Biotechnology Information (http://ncbi/nlm/nih.gov) **[11].** SignalP 4.0 server was employed for detection of signal peptide within the amino acid sequences (http://www.cbs.dtu.dk/services/SignalP/) **[12].** After processing only main chain of peptides were used for further analysis. Detailed information regarding sequences was mined from Protein Information Resources (PIR) konwlwdgebase and literatures **[13].** Protein Information Resources (PIR) is an integrated public bioinformatics resource which helps the genomic and proteomic research. For better understanding a sequence ID code was given to each molecule. Physicochemical characterization including number of amino acids, molecular weight, theoretical isoelectric point (pI), amino acid composition (%) profile, number of positively charged (Arg + Lys) and negatively charged (Asp + Glu) amino acid residues, instability index and Grand Average of Hydropathicity (GRAVY) value were calculated with the help of Expasy ProtParam tool (http://expasy.org/tools/protparam.html).

Discussion:

In the present study, snake venom toxins (cytotoxins and short neurotoxin of *naja annulifera* and *naja naja*) and related non-toxin proteins of other chordates were analyzed with the help of bioinformatical tools **Table 1 (see supplementary material)**. The analysis of amino acid composition of each sequence depicts that conservation of cysteine amino acid took place in different molecules in different organisms **Table 2 (see supplementary material)**. Cytotoxins, short neurotoxins and related non-toxin proteins are similar in their cysteine profile but substantially different in composition of other amino acids. Cysteine profile is conserved because it is responsible for disulphide bridging which is crucial for maintenance of internal core structure of three-finger proteins **[7].** Positively charged lysine amino acid is present in very high percentage in cytotoxin, short neurotoxin and in xenoxins. Lysine with the help of ionic bonds interacts with other charged biomolecules of cells, increasing the reactivity of the protein. Lethality of cytotoxins is facilitated by an invariant lysine residue of these (cytotoxins) peptides **[14].** Short neurotoxin binding to nAChR is governed by positively charged amino acid lysine **[15, 16]**. Arginine is also present in high amount in short neurotoxin which is another positively charged amino acid, is responsible for the receptor binding mechanism. Both Lysine and Arginine and their adequate presence help cytotoxin and short neurotoxin to become an effective lethal bio-molecule. Additionally short neurotoxins also manage negatively charged amino acid efficiently than cytotoxins. Negatively charged aspartic acid and glutamic acid assisted proper attachment to membrane receptor **[17].** Very high amount of negatively charged amino acid is also present in other three-finger proteins like Plethodontid modulating factor (PMF) and Lymphocyte antigen 6H (Ly6H) molecules of different organisms. In PMF the high amount of negativity is contributed by the presence of Aspartic acid and Glutamic acid residues in the sequence.

Three -finger proteins function mainly by binding to other proteins. The PMF also follows that direction by binding to positively charged female receptors for pheromone attachment **[18].** More negativity of PMF by presence of negatively charged amino acid accelerates the binding mechanism of pheromone to a receptor in very expeditiously way. Other non-toxin protein of chordates contains a balanced proportion of positively and negatively charged amino acids. It is because these proteins play different key regulatory cellular processes within the internal physiological system (cellular communication system, complement system and nervous system). Large perturbation in amino acid composition affects the system in a detrimental path, although they evolved efficiently for better adaptation to system **[19]**. Family members of a particular family of three-finger proteins present in different species show same conservation of amino acid composition profile (e.g., Lynx-1). Moderate deviations were also evidenced in complement system proteins (CD59). Involvement in different biological functions is one of the driving forces which contribute the vivid amino acid composition profile of these proteins.

Table 3 (see supplementary material) furnishes details of the physicochemical characterization, which shows that the minimum amino acid residue containing protein is snake venom cytotoxins and maximum amino acid residue containing protein is Lynx-1, a neuromodulator. Computation of Isoelectric point (theoretical pI) and molecular weight (Mw) of an amino acid sequence is worthy because these data dictate the approximate area of a 2D-gel where a protein of interest may be detected. The cytotoxins and short neurotoxins are highly basic in nature (pI 8.69 - 9.48) where as other related protein molecules are acidic or basic. PMF is one of the chief acidic molecules with a pI range of 3.74 to 3.96. Instability index shows that Xenoxins, HLMP1, HEP21 and Lynx-1 are stable in nature (instability index <40). The relative volume of a protein

occupied by its aliphatic side chains is termed as Aliphatic index (AI). Aliphatic index plays role in protein thermal stability. With a high Aliphatic index proteins are more thermally stable. Aliphatic amino acids also are hydrophobic in nature. The aliphatic index of cytotoxins in the range of 66.5 to 84.33 indicated that these proteins are thermally stable as well as they contain high amount of hydrophobic amino acids. Co-presence of hydrophobic and polar (charged) residues within cytotoxins generates amphipathic nature of cytotoxins. For biological membrane perturbation this is an important criterion for a molecule. Short neurotoxin ranges an aliphatic index of 30.33 to 54.26. All different family of three-finger proteins exhibit family specific aliphatic index profile. All proteins included in this study are hydrophilic (negative GRAVY value), whereas exceptionally SLURP1 and SLURP2 are slightly hydrophobic in nature. Short neurotoxin with GRAVY value of -1.213 is the most potent hydrophilic molecule.

Conclusion:

In the post-genomic era not only the generation of data but also proper assimilation of knowledge from these data is a significant deed. Development of different computational resources for exploration of biological data thrusts the discoveries of new insights into the different areas of biological sciences. Comparative physicochemical characterization of proteins from its sequence of a protein superfamily portrays the family specific molecular compositional strategy for improve system adaptability. The present study on snake venom toxin proteins and non-toxin body proteins help to understand what kind of compositional biasness and differences plays role for adaptation to different biological systems namely venom system, pheromone system, complement system and cellular communication system. Notably the exploitation of a protein scaffold which is involved in diverse biological function, used in snakes as venom architecture describes the uniqueness of process of evolution. Physicochemical characterization of these proteins describes within the Laboratory of Nature how proteins are engineered for customized biological needs. This in turn assists to generate therapeutic molecules of medicinal importance.

References:

[1] Calvete JJ *et al. FEBS Lett*. 2009 **583**: 1736 [PMID: 19303875]
[2] Vidal N & Hedges SB, *C R Biol*. 2005 **328**: 1000 [PMID: 16286089]
[3] Fry BG, *Genome Res*. 2005 **15**: 403 [PMID: 15741511]
[4] Chippaux JP, *Bull World Health Organ*. 1998 **76**: 515 [PMID: 9868843]
[5] Brunda G & Sashidhar RB, *Indian J Med Res*. 2007 **125**: 661 [PMID: 17642502]
[6] Phui Yee JS *et al*. *Biochimie*. 2004 **86**: 137 [PMID: 15016453]
[7] Tsetlin V, *Eur J Biochem*. 1999 **264**: 281 [PMID: 10491072]
[8] Dufton MJ & Hider RC, *Pharmacol Ther*. 1988 **36**: 1 [PMID: 3277206]
[9] Changeux JP, *Trends Pharmacol Sci*. 1990 **11**: 485 [PMID: 2080554]
[10] Miwa JM *et al*. *Neuron*. 2006 **51**: 587 [PMID: 16950157]
[11] Sayers EW, *Nucleic Acids Res*. 2012 **40**: D13 [PMID: 22140104]
[12] Petersen TN *et al*. *Nat Methods*. 2011 **8**: 785 [PMID: 21959131]
[13] Barker WC *et al*. *Nucleic Acids Res*. 1999 **27**: 39 [PMID: 9847137]
[14] Jayaraman G *et al*. *Protein Sci*. 2000 **9**: 637 [PMID: 10794406]
[15] Mordvintsev DY *et al*. *Comput Biol Chem*. 2005 **29**: 398 [PMID: 16290328]
[16] Ackermann EJ & Taylor P, *Biochemistry*. 1997 **36**: 12836 [PMID: 9335541]
[17] Lesovoy DM *et al*. *Biophys J*. 2009 **97**: 2089 [PMID: 19804741]
[18] Palmer CA *et al*. *J Mol Evol*. 2010 **70**: 427 [PMID: 20407892]
[19] Dorus S *et al*. *Cell*. 2004 **119**: 1027 [PMID: 15620360]

Edited by P Kangueane

Citation: Panda & Chandra, Bioinformation 8(18): 891-896 (2012)

BIOINFORMATION open access

Supplementary materials:

Table 1: Sequence ID, database accession number, source organism of cytotoxins, short neurotoxins and related non-toxin proteins and their corresponding functions

Sequence ID	Protein Name	Accession no.	Source organism	Function
CX1.BEC	Cytotoxin 1	117664	*Naja annulifera*	Shows cytolytic activity
CX2.BEC	Cytotoxin 2	117676	*Naja annulifera*	Shows cytolytic activity
CX3.BEC	Cytotoxin 3	117696	*Naja annulifera*	Shows cytolytic activity
CX4.BEC	Cytotoxin 4	117714	*Naja annulifera*	Shows cytolytic activity
CX5.BEC	Cytotoxin 5	117718	*Naja annulifera*	Shows cytolytic activity
CX6.BEC	Cytotoxin 6	117722	*Naja annulifera*	Shows cytolytic activity
CX7.BEC	Cytotoxin 7	117723	*Naja annulifera*	Shows cytolytic activity
CX8.BEC	Cytotoxin 8	117724	*Naja annulifera*	Shows cytolytic activity
CX9.BEC	Cytotoxin 9	117725	*Naja annulifera*	Shows cytolytic activity
CX10.BEC	Cytotoxin 10	117660	*Naja annulifera*	Shows cytolytic activity
CX1.IC	Cytotoxin 1	117667	*Naja naja*	Shows cytolytic activity
CX2.IC	Cytotoxin 2	117680	*Naja naja*	Shows cytolytic activity
CX3.IC	Cytotoxin 3	117700	*Naja naja*	Shows cytolytic activity
CX7.IC	Cytotoxin 7	298351639	*Naja naja*	Shows cytolytic activity
NXS1.BEC	Short neurotoxin 1	55977300	*Naja annulifera*	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site.
NXS2.BEC	Short neurotoxin 2	128982	*Naja annulifera*	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site.
NXS3.BEC	Short neurotoxin 3	128986	*Naja annulifera*	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site.
NXS4.BEC	Short neurotoxin 4	128989	*Naja annulifera*	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site.
Xenoxin 1	Xenoxin 1	586258	*Xenopus laevis*	Lacks alpha-neurotoxic activity, channel protein activation
Xenoxin 2	Xenoxin 2	731166	*Xenopus laevis*	Lacks alpha-neurotoxic activity, channel protein activation
Xenoxin 3	Xenoxin 3	731167	*Xenopus laevis*	Lacks alpha-neurotoxic activity, channel protein activation
HLMP 1	Leukocyte membrane protein 1	5714377	*Eptatretus stoutii*	Acts upon complement system
HEP21.C	Hep21 protein	45383131	*Gallus gallus*	Related to Ly-6 protein
HEP21.T	Hep21 protein	326930094	*Meleagris gallopavo*	Related to Ly-6 protein
PMF.PS	Plethodontid modulating factor	113912825	*Plethodon shermani*	Act as a pheromone protein, affects female receptivity
PMF.PC	Plethodontid modulating factor	113913043	*Plethodon cheoah*	Act as a pheromone protein, affects female receptivity
PMF.PY	Plethodontid modulating factor	113913185	*Plethodon yonahlossee*	Act as a pheromone protein, affects female receptivity
CD59.H	CD59 glycoprotein	116021	*Homo sapiens*	Potent inhibitor of the complement membrane attack complex (MAC) action
CD58.M	CD59 glycoprotein	13878360	*Mus musculus*	Potent inhibitor of the complement membrane attack complex (MAC) action
CD59.R	CD59 glycoprotein	2507508	*Rattus norvegicus*	Potent inhibitor of the complement membrane attack complex (MAC) action
Ly6H.H	Lymphocyte antigen 6H	10720070	*Homo sapiens*	Involved in cellular interaction, activation of T lymphocytes
Ly6H.CM	Lymphocyte antigen 6H	167008973	*Macaca fascicularis*	Involved in cellular interaction
Ly6H.B	Lymphocyte antigen 6H	167008972	*Bos taurus*	Involved in cellular interaction
Ly6H.M	Lymphocyte antigen 6H	10720078	*Mus musculus*	Involved in cellular interaction
SLURP1.H	Secreted Ly-6/uPAR-related protein 1	3287957	*Homo sapiens*	Has an antitumor activity, Implicated in maintaining the physiological and structural integrity of the keratinocyte layers of the skin.
SLURP1.M	Secreted Ly-6/uPAR-related protein 1	14916717	*Mus musculus*	T cell activation & cell to cell adhesion, Was found to be a marker of late differentiation of the skin
SLURP2.H	Secreted Ly6/uPAR related protein 2	74727391	*Homo sapiens*	Regulation of lymphocyte function
SLURP2.M	Secreted Ly6/uPAR related protein 2	123778205	*Mus musculus*	Regulation of lymphocyte function
Lynx1.H	Ly-6/neurotoxin-like protein 1	47117907	*Homo sapiens*	Seems to modulate nicotinic acetylcholine receptors
Lynx1.C	Ly-6/neurotoxin-like protein 1	61214436	*Pan troglodytes*	Seems to modulate nicotinic acetylcholine receptors
Lynx1.RM	Ly-6/neurotoxin-like protein 1	46576878	*Macaca mulatta*	Seems to modulate nicotinic acetylcholine receptors
Lynx1.BM	Ly-6/neurotoxin-like protein 1	75040497	*Saimiri boliviensis*	Seems to modulate nicotinic acetylcholine receptors
Lynx1.B	Ly-6/neurotoxin-like protein 1	126256577	*Bos taurus*	Seems to modulate nicotinic acetylcholine receptors
Lynx1.M	Ly-6/neurotoxin-like protein 1	24212024	*Mus musculus*	Seems to modulate nicotinic acetylcholine receptors

Table 2: Amino acid composition profile (in %) of various snake venom toxin proteins and related non-toxin proteins of other chordates

Seq. ID	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
CX1.BEC	1.7	13.3	3.3	1.7	1.7	3.3	1.7	1.7	15.0	6.7	3.3	5.0	8.3	0.0	1.7	6.7	6.7	13.3	1.7	3.3
CX2.BEC	3.3	13.3	5.0	1.7	1.7	3.3	1.7	1.7	15.0	8.3	6.7	5.0	10.0	0.0	1.7	1.7	5.0	8.3	1.7	5.0
CX3.BEC	1.7	13.3	3.3	1.7	1.7	3.3	0.0	1.7	15.0	8.3	3.3	6.7	8.3	0.0	1.7	3.3	6.7	11.7	1.7	6.7
CX4.BEC	1.7	13.3	1.7	0.0	1.7	3.3	0.0	3.3	16.7	8.3	3.3	10.0	8.3	0.0	1.7	3.3	6.7	10.0	1.7	5.0
CX5.BEC	3.3	13.3	1.7	1.7	1.7	3.3	1.7	3.3	15.0	6.7	6.7	8.3	10.0	0.0	1.7	1.7	5.0	8.3	1.7	5.0
CX6.BEC	3.3	13.3	3.3	1.7	1.7	3.3	1.7	1.7	15.0	8.3	6.7	6.7	10.0	0.0	1.7	1.7	5.0	8.3	1.7	5.0
CX7.BEC	3.3	13.3	1.7	1.7	1.7	3.3	1.7	1.7	15.0	8.3	6.7	8.3	10.0	0.0	1.7	1.7	5.0	8.3	1.7	5.0
CX8.BEC	1.7	13.3	1.7	1.7	1.7	3.3	1.7	1.7	15.0	8.3	3.3	8.3	8.3	0.0	1.7	3.3	6.7	11.7	1.7	5.0
CX9.BEC	3.3	13.3	5.0	3.3	1.7	3.3	1.7	3.3	11.7	6.7	3.3	6.7	6.7	0.0	3.3	5.0	6.7	11.7	0.0	3.3
CX10.BEC	3.3	13.3	3.3	3.3	1.7	3.3	1.7	5.0	10.0	6.7	3.3	8.3	6.7	1.7	3.3	5.0	6.7	10.0	0.0	3.3
CX1.IC	3.3	13.3	3.3	1.7	0.0	3.3	0.0	3.3	15.0	10.0	3.3	10.0	6.7	0.0	3.3	3.3	5.0	8.3	0.0	6.7
CX2.IC	3.3	13.3	3.3	0.0	1.7	3.3	0.0	1.7	15.0	10.0	3.3	6.7	8.3	0.0	3.3	3.3	5.0	11.7	0.0	6.7
CX3.IC	3.3	13.3	3.3	0.0	1.7	3.3	0.0	3.3	15.0	10.0	3.3	10.0	6.7	0.0	3.3	3.3	5.0	10.0	0.0	5.0
CX7.IC	3.3	13.3	5.0	1.7	0.0	3.3	0.0	3.3	15.0	10.0	3.3	8.3	6.7	0.0	3.3	3.3	5.0	8.3	0.0	6.7
NXS1.BEC	0.0	13.1	3.3	6.6	0.0	8.2	3.3	4.9	9.8	1.6	0.0	8.2	6.6	4.9	6.6	6.6	11.5	1.6	1.6	1.6
NXS2.BEC	0.0	13.1	1.6	3.3	0.0	9.8	3.3	8.2	11.5	0.0	1.6	6.6	6.6	4.9	8.2	4.9	8.2	3.3	1.6	3.3
NXS3.BEC	0.0	13.1	1.6	4.9	1.6	9.8	1.6	9.8	11.5	1.6	1.6	4.9	4.9	6.6	4.9	4.9	8.2	3.3	1.6	3.3
NXS4.BEC	0.0	13.1	1.6	4.9	1.6	9.8	1.6	9.8	11.5	1.6	1.6	4.9	4.9	4.9	6.6	4.9	8.2	3.3	1.6	3.3
Xenoxin-1	6.1	12.1	1.5	6.1	1.5	6.1	0.0	4.5	12.1	10.6	6.1	6.1	1.5	4.5	3.0	4.5	12.1	1.5	0.0	0.0
Xenoxin-2	6.1	12.1	3.0	6.1	1.5	4.5	0.0	6.1	15.2	10.6	6.1	7.6	1.5	3.0	3.0	4.5	7.6	1.5	0.0	0.0
Xenoxin-3	7.6	12.1	3.0	6.1	1.5	4.5	0.0	4.5	13.6	10.6	4.5	6.1	1.5	4.5	3.0	4.5	9.1	3.0	0.0	0.0
HLMP1	4.1	10.8	6.8	5.4	0.0	5.4	1.4	2.7	17.6	4.1	1.4	5.4	1.4	5.4	1.4	6.8	10.8	8.1	0.0	1.4
Hep21.C	5.7	11.4	6.8	8.0	1.1	5.7	1.1	3.4	6.8	6.8	0.0	3.4	2.3	3.4	8.0	6.8	9.1	3.4	1.1	5.7
Hep21.T	5.8	11.6	4.7	10.5	1.2	5.8	0.0	3.5	7.0	7.0	1.2	3.5	2.3	3.5	7.0	7.0	8.1	2.3	1.2	7.0
PMF.PS	3.0	12.1	12.1	16.7	4.5	9.1	1.5	3.0	3.0	6.1	3.0	7.6	3.0	1.5	1.5	0.0	6.1	1.5	0.0	4.5
PMF.PC	3.0	12.1	12.1	16.7	4.5	9.1	1.5	3.0	4.5	6.1	3.0	7.6	3.0	1.5	0.0	0.0	6.1	1.5	0.0	4.5
PMF.PY	3.0	12.1	12.1	16.7	3.0	9.1	3.0	1.5	6.1	6.1	3.0	4.5	3.0	1.5	0.0	0.0	6.1	3.0	0.0	6.1
CD59.H	5.2	13.0	6.5	6.5	5.2	1.3	1.3	1.3	7.8	9.1	0.0	13.0	2.6	3.9	2.6	2.6	7.8	3.9	1.3	5.2
CD59.M	2.7	13.7	5.5	4.1	4.1	2.7	2.7	2.7	5.5	6.8	4.1	5.5	2.7	9.6	2.7	9.6	4.1	5.5	1.4	4.1
CD59.R	6.3	12.7	6.3	2.5	3.8	1.3	0.0	2.5	6.3	8.9	0.0	10.1	3.8	6.3	5.1	10.1	2.5	6.3	1.3	3.8
LY6H.H	4.4	11.1	10.0	1.1	4.4	4.4	3.3	3.3	6.7	5.6	2.2	4.4	3.3	3.3	3.3	13.3	6.7	6.7	1.1	1.1
LY6H.CM	3.3	10.0	10.0	1.1	4.4	4.4	3.3	3.3	6.7	5.6	2.2	4.4	3.3	3.3	3.3	13.3	6.7	7.8	1.1	2.2
LY6H.B	3.3	11.1	10.0	1.1	4.4	3.3	4.4	3.3	6.7	5.6	2.2	4.4	3.3	3.3	2.2	12.2	7.8	7.8	2.2	1.1
LY6H.M	4.7	11.8	10.6	1.2	4.7	2.4	3.5	3.5	7.1	5.9	2.4	4.7	3.5	3.5	3.5	12.9	4.7	7.1	1.2	1.2
SLURP1.H	7.4	12.3	4.9	6.2	3.7	1.2	1.2	3.7	3.7	6.2	2.5	2.5	6.2	1.2	4.9	11.1	12.3	6.2	0.0	2.5
SLURP1.M	9.1	11.4	4.5	4.5	8.0	5.7	2.3	3.4	3.4	4.5	2.3	4.5	6.8	2.3	3.4	8.0	8.0	6.8	0.0	1.1
SLURP2.H	2.7	13.3	5.3	1.3	1.3	9.3	6.7	5.3	1.3	9.3	1.3	2.7	5.3	2.7	4.0	9.3	10.7	5.3	1.3	1.3
SLURP2.M	2.7	13.3	4.0	0.0	2.7	8.0	4.0	5.3	2.7	9.3	1.3	2.7	8.0	2.7	4.0	16.0	5.3	5.3	1.3	1.3
Lynx1.H	5.5	13.7	5.5	1.4	2.7	4.1	2.7	0.0	4.1	2.7	5.5	5.5	5.5	1.4	5.5	6.8	11.0	6.8	0.0	9.6
Lynx1.C	5.5	13.7	5.5	1.4	2.7	4.1	2.7	0.0	4.1	2.7	5.5	5.5	5.5	1.4	5.5	6.8	11.0	6.8	0.0	9.6
Lynx1.RM	5.5	13.7	5.5	1.4	2.7	2.7	2.7	0.0	4.1	2.7	5.5	5.5	5.5	1.4	4.1	9.6	11.0	6.8	0.0	9.6
Lynx1.BM	5.5	13.7	5.5	1.4	2.7	4.1	2.7	0.0	4.1	2.7	5.5	5.5	5.5	1.4	4.1	8.2	11.0	6.8	0.0	9.6
Lynx1.B	4.1	13.7	4.1	2.7	2.7	4.1	2.7	0.0	4.1	2.7	5.5	5.5	5.5	1.4	4.1	9.6	11.0	6.8	0.0	9.6
Lynx1.M	6.8	13.7	2.7	2.7	4.1	4.1	2.7	0.0	5.5	2.7	5.5	2.7	5.5	4.1	5.5	6.8	9.6	5.5	0.0	11.0

Table 3: Physicochemical characterization of different toxin and non-toxin protein sequences

Sequence ID	No. of amino acids	Molecular weight	pI	R^-	R^+	Instability index	Aliphatic Index	GRAVY
CX1.BEC	60	6696.1	9.15	3	10	47.71	72.83	-0.007
CX2.BEC	60	6858.3	8.99	4	10	66.54	66.5	-0.115
CX3.BEC	60	6839.2	9.11	3	10	47.89	74.5	-0.035
CX4.BEC	60	6802.3	9.48	1	11	46.69	76.17	-0.073
CX5.BEC	60	6856.4	9.26	2	10	62.5	66.5	-0.103
CX6.BEC	60	6857.3	9.13	3	10	69.4	66.5	-0.115
CX7.BEC	60	6856.4	9.26	2	10	69.4	66.5	-0.115
CX8.BEC	60	6812.2	9.26	2	10	60.12	74.5	-0.067
CX9.BEC	60	6668.9	8.69	5	9	48.78	76.17	-0.037
CX10.BEC	60	6681.9	8.7	4	8	38.09	77.83	-0.025
CX1.IC	60	6791.2	9.24	3	11	51.27	79.5	-0.192
CX2.IC	60	6763.2	9.36	2	11	52.18	82.67	0.068
CX3.IC	60	6745.2	9.38	2	11	33.94	84.33	0.005
CX7.IC	60	6792.2	9.11	4	11	52.21	79.5	-0.192
NXS1.BEC	61	6843.6	8.71	6	10	79.01	30.33	-1.213
NXS2.BEC	61	6915	9.46	3	12	59.26	41.48	-0.928
NXS3.BEC	61	6885	9.03	4	10	56.6	54.26	-0.577
NXS4.BEC	61	6913.1	9.18	4	11	53.45	54.26	-0.593
Xenoxin 1	66	7235.6	8.88	5	10	44.73	69.55	-0.174
Xenoxin 2	66	7345.8	9.02	6	12	40.3	75.45	-0.239
Xenoxin 3	66	7258.6	8.87	6	11	34.12	75.45	-0.197
HLMP 1	74	8101.3	8.82	9	14	7.97	53.92	-0.799
HEP21.C	88	10001.2	6.73	13	13	26.4	55.45	-0.703
HEP21.T	86	9830.1	5.44	13	12	37.7	53.37	-0.664
PMF.PS	66	7487	3.74	19	3	70	42.88	-0.774
PMF.PC	66	7459	3.74	19	3	71.14	42.88	-0.765
PMF.PY	66	7498.1	3.96	19	4	69.09	41.36	-0.833
CD59.H	77	8961.1	5.18	10	8	33.78	57.01	-0.578
CD58.M	73	8412.6	6.04	7	6	61.92	56.03	-0.315
CD59.R	79	8936.1	8.09	7	9	45.14	69.11	-0.333
Ly6H.H	90	9860.1	6.28	10	9	44.63	58.44	-0.261
Ly6H.CM	90	9948.2	6.28	10	9	44.73	60.56	-0.277
Ly6H.B	90	10012.3	6.02	10	8	47.57	60.56	-0.224
Ly6H.M	85	9456.7	6.28	10	9	47.36	61.88	-0.241
SLURP1.H	81	8853.1	5.16	9	7	56.86	63.83	0.017
SLURP1.M	88	9462.8	5.48	8	6	58.1	59.89	0.114
SLURP2.H	75	8023.2	6.53	5	4	53.41	75.33	0.096
SLURP2.M	75	7948.2	8.12	3	5	68.78	75.33	0.207
Lynx1.H	73	8278.4	8.09	5	7	27.8	36.03	-0.321
Lynx1.C	73	8278.4	8.09	5	7	27.8	36.03	-0.321
Lynx1.RM	73	8239.4	7.64	5	6	37.99	36.03	-0.275
Lynx1.BM	73	8209.3	7.64	5	6	35.94	36.03	-0.27
Lynx1.B	73	8239.4	7.64	5	6	34.9	34.66	-0.305
Lynx1.M	73	8372.6	8.56	4	8	36.55	33.42	-0.281