

Sequence analysis and phylogenetic study of some toxin proteins of snakes and related non-toxin proteins of chordates

Subhamay Panda¹ & Goutam Chandra^{2*}

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

Received February 10, 2013; Accepted February 12, 2013; Published March 02, 2013

Abstract:

Snakes are equipped with their venomous armory to tackle different prey and predators in adverse natural world. The venomous composition of snakes is a mix of biologically active proteins and polypeptides. Among different components snake venom cytotoxins and short neurotoxin are non-enzymatic polypeptide candidates with in the venom. These two components structurally resembled to three-finger protein superfamily specific scaffold. Different non-toxin family members of three-finger protein superfamily are involved in different biological roles. In the present study we analyzed the snake venom cytotoxins, short neurotoxins and related non-toxin proteins of different chordates in terms of amino acid sequence level diversification profile, polarity profile of amino acid sequences, conserved pattern of amino acids and phylogenetic relationship of these toxin and non-toxin protein sequences. Sequence alignment analysis demonstrates the polarity specific molecular enrichment strategy for better system adaptivity. Occurrence of amino acid substitution is high in number in toxin sequences. In non-toxin body proteins there are less amino acid substitutions. With the help of conserved residues these proteins maintain the three-finger protein scaffold. Due to system specific adaptation toxin and non-toxin proteins exhibit a varied type of amino acid residue distribution in sequence stretch. Understanding of Natural invention scheme (recruitment of venom proteins from normal body proteins) may help us to develop futuristic engineered bio-molecules with remedial properties.

Keywords: Snake venom cytotoxin, Short neurotoxins, Non-toxin proteins of chordates, Amino acid, Sequence level diversification, Polarity profile, Phylogenetic study, Evolutionary process.

Background:

In the domain of biology and biochemistry interpreting the evolutionary process of life is of great importance. Study of molecular level evolution of proteins helps us to understand the evolution of life. The journey of evolution assisted with natural selection generates improved biologically functional protein molecules [1]. Different organisms use different adaptive strategy for existence in the adverse natural world. From ancient times mankind was enchanted by the snakes and their astonishing features. The process of evolution gifted a precious thesaurus to the snakes, i.e., their venom. The primary function of snake venom is to incapacitation and immobilization of the prey of the snakes (as an offensive

armory). Evolved snake venom to aid in catching prey exhibits fatal and enfeebling effect. The secondary function of venom is to serve as defensive machinery against their predators. Snake venom also assists in digestion of variety of diets of snakes [2]. The evolution of venoms favors the survival of snakes in different environment [3]. Snake venom is a cornucopia of biologically active polypeptides and non-polypeptide constituents [4]. Among the different venomous component snake venom cytotoxins and short neurotoxins are non-enzymatic polypeptide candidates with a molecular weight of 5-10 kDA [4]. In the cobra venom, fifty percent of the dry weight accounts for its cytotoxin components [5]. The characteristics of cytotoxins and short neurotoxins are i) three

β -strand loops ii) small globular hydrophobic core iii) four conserved disulfide bridges and iv) 60-62 amino acid residues [6, 7, 8]. The lethality of cytotoxin is due to its pore formation capacity into the biological membranes [9]. Short neurotoxins exert their effect by blocking the neuromuscular transmission through selective binding to muscle nicotinic acetyl-choline receptors (nAChR) [10]. Cytotoxin and short neurotoxin are enriched with disulfide bonds. As a family member of 'Three-finger protein' superfamily, cytotoxins and short neurotoxins exhibit 'three-finger' appearance which is a chief structural characteristics feature of this superfamily. The name of three-finger protein is for three loops of three-finger proteins that are outstretched from the core region of the protein which is looked like stretch out three fingers of a hand [6, 8]. For proper structural maintenance three-finger proteins are rich in disulfide bonds [8]. Three-finger protein scaffold is also present in different chordates where they employed in diverse biological functions. Plethodontid modulating factor (PMF), Lymphocyte antigen 6H (Ly6H), CD59 glycoprotein (CD59), Ly-6/neurotoxin-like protein 1 (Lynx-1) are different non-toxin family members of three-finger protein superfamily.

These non-toxin family members are involved in diverse physiological system namely pheromone system, complement system, cellular communication system and central nervous system etc [8, 11, 12]. Snake venom proteins recruited into venom proteome by a process of evolution in which genes of normal body proteins that are engaged in key regulatory process within the body are duplicated. Selective expression of these duplicated genes in the venom gland produces deadly cocktail of lethal toxin molecules [12]. Previously a study on the physicochemical characterization and functional analysis of snake venom toxin proteins & related proteins (non-toxin) of chordates was done to understand their compositional differences, physicochemical properties and functions in a comparative manner [13]. The aim of the present study was to analyze the diversification profile of amino acid sequences, polarity profile of amino acid sequences, conservation pattern of amino acid residues and phylogenetic relationship of snake venom toxin proteins and related non-toxin three-finger proteins of different chordates like hagfish, bird, frog, mouse, rat, human etc.

Methodology:

Protein sequences of cytotoxins, short neurotoxins and related non-toxin body proteins were retrieved from National Centre for Biotechnology Information (<http://ncbi.nlm.nih.gov>) [14]. Presence of signal peptide within the amino acid sequences and cleavage sites of signal peptides was analyzed using SignalP 4.0 algorithm which is a neural network-based method (<http://www.cbs.dtu.dk/services/SignalP/>) [15]. Main chain of protein sequences were selected for further analysis. For simplicity a sequence ID code was given to each molecule. Sequence related information's was mined from Protein Information Resources (PIR) knowledgebase [16]. Multiple sequence alignment is done with Clustal-X program, followed by manual inspection for errors [17]. Aligned sequence sets were represented in Clustal-X coloring scheme and polarity coloring scheme using Clustal-X and Jalview tool respectively [17, 18].

Sequence logos were generated using WebLogo application, which helps to dictate the sequence conservation and relative

frequencies of the amino acid residues at each position [19, 20]. Family specific conserved pattern of amino acid residues in toxin and non-toxin proteins were extracted using PRATT version 2.1 [21]. Phylogenetic trees were re-constructed using neighbor-joining method implemented in MEGA 4.0 version [22]. For Bayesian inference of phylogeny, MrBayes 3.1.2 was used [23].

Results & Discussion:

In this work, various snake venom cytotoxins and short neurotoxins (*naja annulifera* and *naja naja*) and related chordate specific non-toxin proteins were analyzed with the help of different bioinformatical packages to address evolutionary process of these molecules **Table 1** (see **supplementary material**). The number of amino acids of main chain of toxin proteins and non-toxin proteins comprise of 60 to 90 amino acid residues (**Table 1**). Multiple sequence alignment of protein sequences (represented in polarity coloring schemes) shows variable sites as well as conserved sites of these proteins which demonstrate the sequence enrichment strategy of these sequences for adaptation to different physiological systems (**Figure 1**). First residue of all sequences is a Leu (hydrophobic amino acid) which is conserved in all peptides except in short neurotoxins (replaced by Met), Secreted Ly-6/uPAR-related protein 1 (SLURP1) of human (replaced by Phe) and Secreted Ly-6/uPAR-related protein 2 (SLURP2) of human (replaced by Ile).

In terms of polarity Leu, Met, Phe, Ile all were hydrophobic in nature. These replacements of residues (non-synonymous substitution) with a same polarity residues hint conservation strategy at physicochemical characteristics level. Cysteine is conserved in total seven sites, which act like an anchoring residue for maintenance of the three-finger fold of the three-finger protein superfamily (**Supplement Figure 1**). After last Cysteine the successive amino acid is Asn which is conserved in all sequences. These conserved residues help to create the signature residues of three-finger protein domain. Substitutions (point mutations) are high in number in toxin specific sequences. Polarity shuffling and hydrophobic (aliphatic) amino acid substitution are very prevalent in cytotoxins and short neurotoxin sequences (**Figure 1**).

In other related non-toxin proteins, substitutions of amino acids at family level are slightly low in number. In xenoxin family the observed substitution of amino acid in terms of polarity is from a negative amino acid (Asp) to a hydrophobic amino acid (Gly) and from a hydrophobic amino acid (Ile) to polar amino acid (Thr). Presence of more amount of positively charged amino acid residue than negatively charged amino acid is observed throughout in xenoxin sequence stretch. In Hep21 protein (HEP21) sequences negatively charged amino acid is equal in number to positively charged amino acid. In PMF the presence of negatively charged amino acid is greater in number than positively charged amino acid. Very versatile amino acid substitution profile is observed within the family members of CD59. Within family members and at superfamily level CD59 has shown a diversified sequence profile. Low level of amino acid sequence conservation of CD59 molecules of different species result in vivid amino acid distribution profile of CD59 [24]. In Ly6H and Lynx-1 amino acids substitution are uncommon and these protein sequences are very conservative in nature within their corresponding family (**Figure 1**).

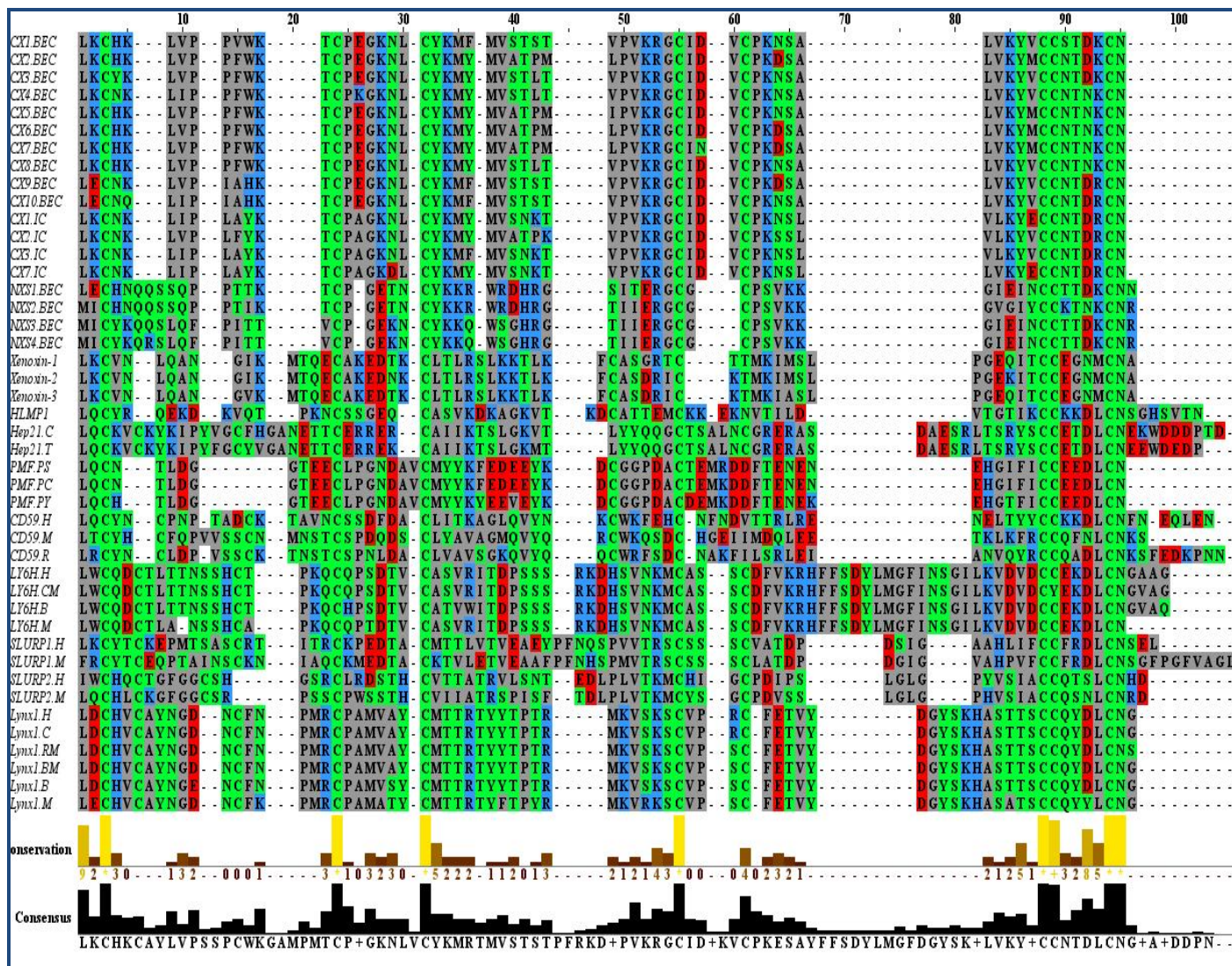


Figure 1: Multiple sequence alignment of various snake venom toxin proteins and related non-toxin proteins of other chordates (polarity coloring scheme).

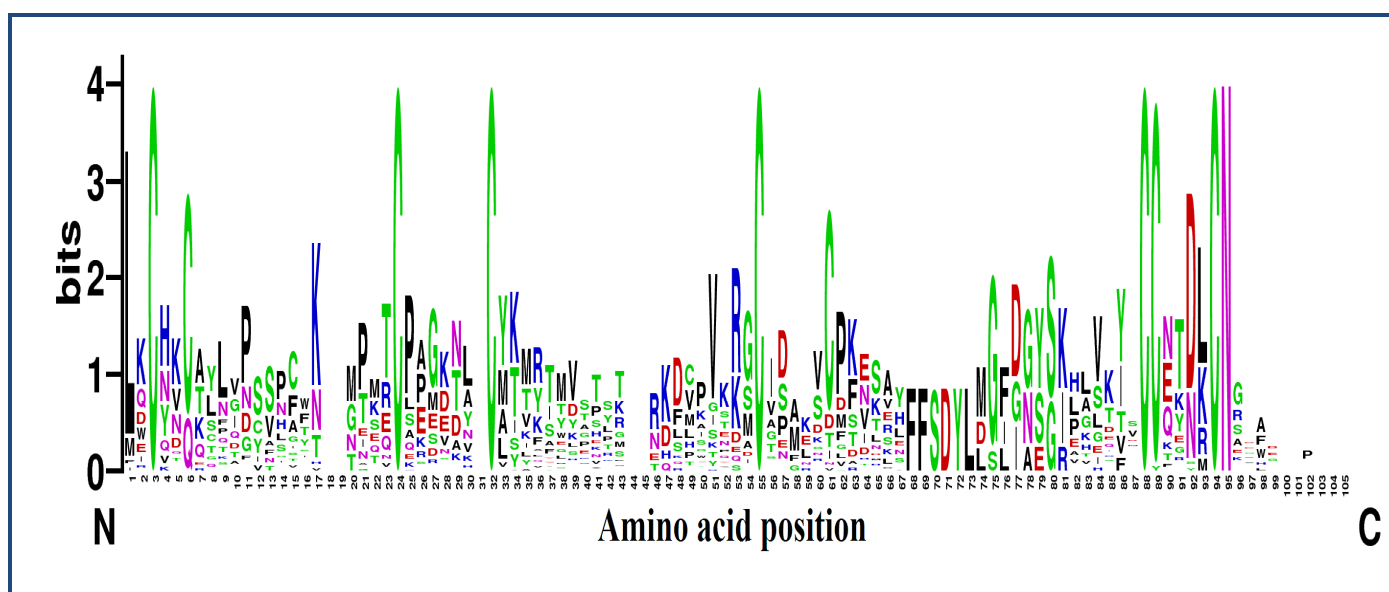


Figure 2: Sequence logo representation of snake venom toxin proteins and related non-toxin proteins of other chordates. Sequence logo was created by WebLogo program (<http://weblogo.berkeley.edu/>).

WebLogo program is used to create sequence logos. With the help of Sequence logos the pattern in sequence conservation has been addressed in a graphical form (Figure 2). Stack of letters in sequence logos represents the amino acid position. The stack height (measured in bits) signifies the degree of sequence conservation at that position and the height of each letter depicts the relative frequency of the corresponding amino acid [19, 20]. Multiple sequence alignment (represented as sequence logo) of cytotoxins and short neurotoxins reveals that despite of the difference in their mode of mechanism of actions i.e., cytotoxin acts upon cell membrane and short-neurotoxin exert action upon nAChR, 24 sites were conserved in their sequences [9, 10]. Eighteen sites (amino acid residue) are identical in nature and amino acid residues of 6 sites are physicochemically conserved (non-synonymous substitution) within the sequences (Supplement Figure 2). Sequence logo of toxin and non-toxin proteins exhibits the pattern of sequence conservation in toxin and non-toxin sequences of three-finger protein superfamily (Figure 2). It is noteworthy that the conserved cysteine profile of all the sequences of toxin and non-toxin proteins indicates their common evolutionary ancestry. With the help of these cysteines the three-finger fold is well maintained in three-finger proteins [8]. Variable sites signify the diversification process of evolution where a single scaffold is used in different system with distinct roles. Individual family specific conserved pattern in protein sequences were tabulated in Table 2 (see supplementary material).

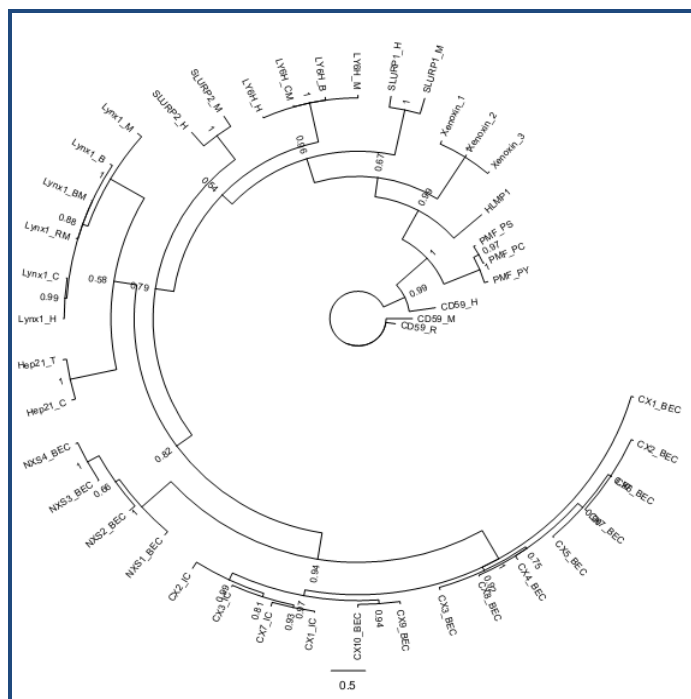


Figure 3: Bayesian phylogenetic analysis of various snake venom toxin proteins and related non-toxin proteins of other chordates.

Phylogenetic analysis of toxin and non-toxin protein sequences using depicts the evolutionary relationship among cytotoxins, short neurotoxins and related non-toxin proteins of other chordates (Supplement Figure 3). Bayesian phylogenetic analysis shows that cytotoxins and short neurotoxins are entirely separate entities in terms of position in phylogenetic tree (Figure 3). Cytotoxin and short neurotoxin belong to two

separate groups but share a common branch point. Species specific distribution of toxin molecules in phylogenetic tree was observed (i.e., cytotoxins of *naja annulifare* and *naja naja* are not mixed within each cluster). Related non-toxin proteins form an entirely separate group of non-toxin body proteins. This can logically be interpreted that Lynx1 and Hep21 share a common branch point with toxin proteins which reasonably logical because cytotoxins and short neurotoxins are originated from an ancestral Lynx1 like molecule [12]. All other non-toxin protein molecules i.e., HEP21, SLURP1, SLURP2, Ly6H, Xenoxins, HMLP1, PMF and CD59 form separate clusters with their corresponding family members.

Conclusion:

Different results signify that cytotoxins and short neurotoxins, which are two important components of snake venom, are originated from simply body proteins enriched with different sequence specific substitution strategy for biological needs. Efficient utilization of hydrophobic amino acids, positively charged and negatively charged amino acids and their distribution profile in toxin sequence make them a tailored killer element in snake venom. An enhanced amphipathic nature added extra advantage to cytotoxins for exerting harmful action upon biological membranes. Variations in physicochemical properties of amino acid within the toxin sequences gave additional opportunity for generation of improved potent toxins. PMF is also efficient to use negatively charged amino acids in sequence stretch for better receptor attachment. Even distribution of negatively charged amino acids at regular intervals helps PMF to provide better chance to attach pheromone receptor in a specific manner. CD59 comprises of vivid amino acid residue distribution in their sequences. Very conserved profile of amino acid sequences in Ly6H and Lynx1 is for their involvement in two important systems i.e., cellular communication system and central nervous system. Within the nature these two systems are very critical for any life form. Distribution of different amino acid residues along the sequence length of three-finger proteins is customized for their adapted biological functional needs. With conserved cysteine these proteins maintain the structural scaffold but at the same time variation of amino acid residues in other parts of sequence facilitates different system dependent necessities. Comparative sequence specific analysis of protein sequences demonstrates how proteins are generated within the nature's testing ground for tailor-made biologic needs. Tracing the Natural protein engineering scheme of three-finger proteins enrich our knowledge which in turn helps to generate bio-molecules with remedial properties.

Acknowledgement:

We are very thankful to Prof. Debesh Chandra Majumder, Chairman, Trinity Trust, Asansol, West Bengal, Prof. Kalyan Kumar Sen, Principal, Gupta College of Technological Sciences, Asansol, West Bengal and The Head, Department of Zoology, The University of Burdwan, Burdwan, West Bengal, for providing infrastructure facilities for carrying out the research work.

References:

- [1] Iida Y, *J Mol Biol.* 1982 **159**: 167 [PMID: 6813507]
- [2] Kang TS *et al. FEBS J.* 2011 **278**: 4544 [PMID: 21470368]
- [3] Panfoli I *et al. Toxins.* 2010 **2**: 417 [PMID: 22069593]

- [4] Phui Yee JS *et al. Biochimie*. 2004 **86**: 137 [PMID: 15016453]
[5] Dubinnyĭ MA *et al. Bioorg Khim*. 2001 **27**: 102 [PMID: 11357394]
[6] Kini RM & Doley R, *Toxicon*. 2010 **56**: 855 [PMID: 20670641]
[7] Ménez A, *Toxicon*. 1998 **36**: 1557 [PMID: 9792172]
[8] Tsetlin V, *Eur J Biochem*. 1999 **264**: 281 [PMID: 10491072]
[9] Dufton MJ & Hider RC, *Pharmacol Ther*. 1988 **36**: 1 [PMID: 3277206]
[10] Changeux JP, *Trends Pharmacol Sci*. 1990 **11**: 485 [PMID: 2080554]
[11] Palmer CA *et al. J Mol Evol*. 2010 **70**: 427 [PMID: 20407892]
[12] Fry BG, *Genome Res*. 2005 **15**: 403 [PMID: 15741511]
[13] Panda S & Chandra G, *Bioinformatics*. 2012 **8**: 891 [PMID: 23144546]
[14] Sayers EW *et al. Nucleic Acids Res*. 2012 **40**: D13 [PMID: 22140104]
[15] Petersen TN *et al. Nat Methods*. 2011 **8**: 785 [PMID: 21959131]
[16] Barker WC *et al. Nucleic Acids Res*. 1999 **27**: 39 [PMID: 9847137]
[17] Thompson JD *et al. Nucleic Acids Res*. 1997 **25**: 4876 [PMID: 9396791]
[18] Waterhouse AM *et al. Bioinformatics*. 2009 **25**: 1189 [PMID: 19151095]
[19] Schneider TD & Stephens RM, *Nucleic Acids Res*. 1990 **18** : 6097 [PMID: 2172928]
[20] Crooks GE *et al. Genome Res*. 2004 **14**: 1188 [PMID: 15173120]
[21] Jonassen I, *Comput Appl Biosci*. 1997 **13**: 509 [PMID: 9367124]
[22] Tamura K *et al. Mol Biol Evol*. 2007 **24**: 1596 [PMID: 17488738]
[23] Ronquist F & Huelsenbeck JP, *Bioinformatics*. 2003 **19**: 1572 [PMID: 12912839]
[24] Rushmere NK *et al. Biochem J*. 1994 **304**: 595 [PMID: 7528012]

Edited by P Kanguane

Citation: Panda & Chandra, *Bioinformatics* 9(5): 259-266 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

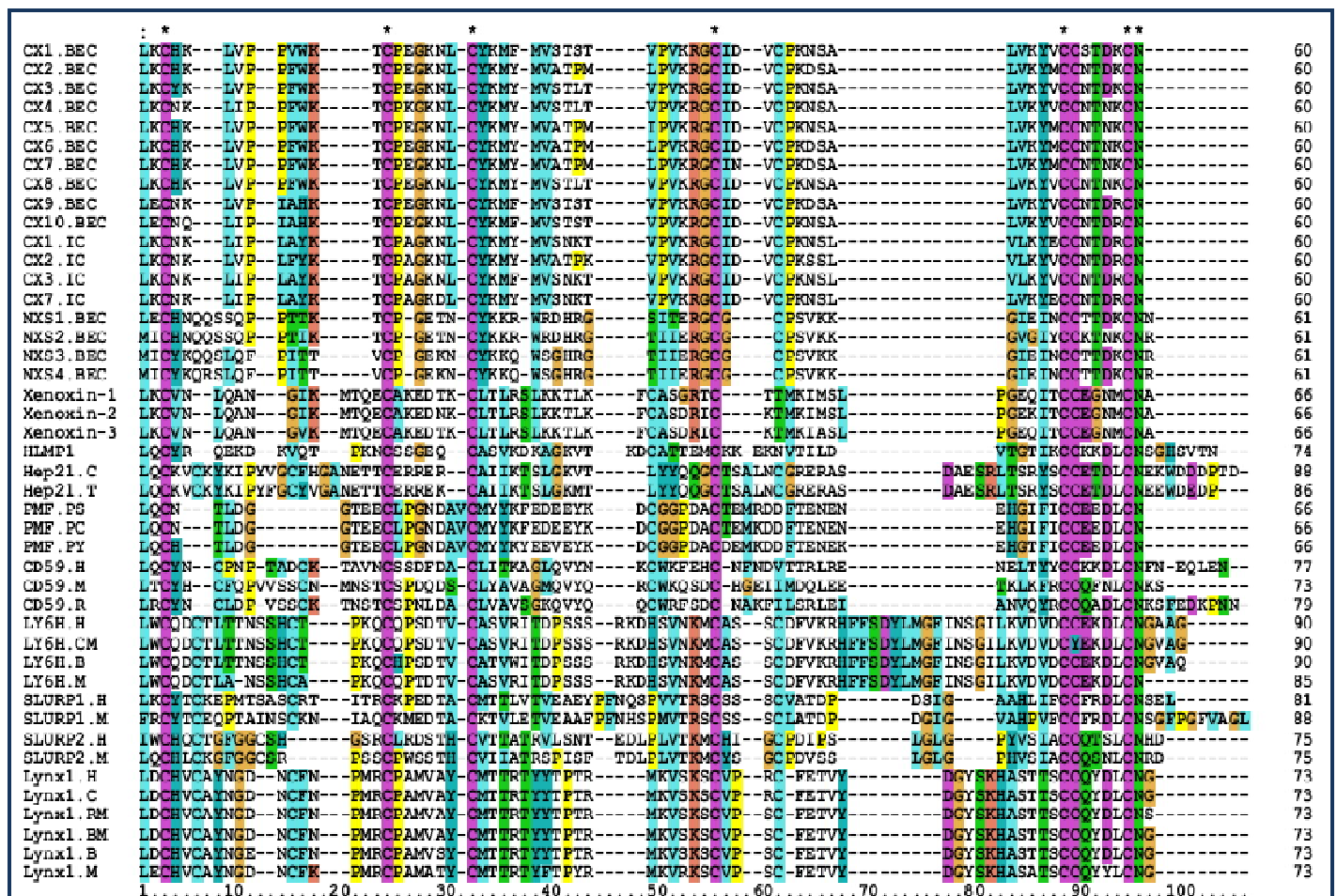
Supplementary material:

Table 1: Sequence ID, protein name, database accession number, source organism, number of amino acid of cytotoxins, short-neurotoxins and related non-toxin proteins.

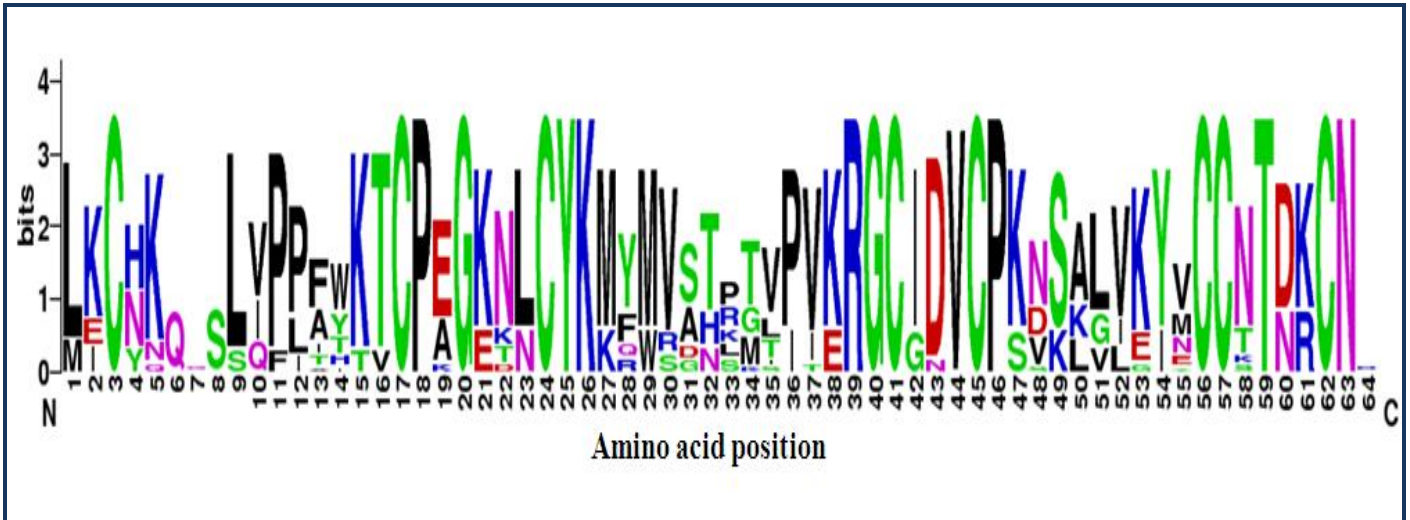
Sequence ID	Protein Name	Accession no.	Source organism	Number of Amino acids
CX1.BEC	Cytotoxin 1	117664	<i>Naja annulifera</i>	60
CX2.BEC	Cytotoxin 2	117676	<i>Naja annulifera</i>	60
CX3.BEC	Cytotoxin 3	117696	<i>Naja annulifera</i>	60
CX4.BEC	Cytotoxin 4	117714	<i>Naja annulifera</i>	60
CX5.BEC	Cytotoxin 5	117718	<i>Naja annulifera</i>	60
CX6.BEC	Cytotoxin 6	117722	<i>Naja annulifera</i>	60
CX7.BEC	Cytotoxin 7	117723	<i>Naja annulifera</i>	60
CX8.BEC	Cytotoxin 8	117724	<i>Naja annulifera</i>	60
CX9.BEC	Cytotoxin 9	117725	<i>Naja annulifera</i>	60
CX10.BEC	Cytotoxin 10	117660	<i>Naja annulifera</i>	60
CX1.IC	Cytotoxin 1	117667	<i>Naja naja</i>	60
CX2.IC	Cytotoxin 2	117680	<i>Naja naja</i>	60
CX3.IC	Cytotoxin 3	117700	<i>Naja naja</i>	60
CX7.IC	Cytotoxin 7	298351639	<i>Naja naja</i>	60
NXS1.BEC	Short neurotoxin 1	55977300	<i>Naja annulifera</i>	61
NXS2.BEC	Short neurotoxin 2	128982	<i>Naja annulifera</i>	61
NXS3.BEC	Short neurotoxin 3	128986	<i>Naja annulifera</i>	61
NXS4.BEC	Short neurotoxin 4	128989	<i>Naja annulifera</i>	61
Xenoxin 1	Xenoxin 1	586258	<i>Xenopus laevis</i>	66
Xenoxin 2	Xenoxin 2	731166	<i>Xenopus laevis</i>	66
Xenoxin 3	Xenoxin 3	731167	<i>Xenopus laevis</i>	66
HLMP 1	Leukocyte membrane protein 1	5714377	<i>Eptatretus stoutii</i>	74
HEP21.C	Hep21 protein	45383131	<i>Gallus gallus</i>	88
HEP21.T	Hep21 protein	326930094	<i>Meleagris gallopavo</i>	86
PMF.PS	Plethodontid modulating factor	113912825	<i>Plethodon shermani</i>	66
PMF.PC	Plethodontid modulating factor	113913043	<i>Plethodon cheoah</i>	66
PMF.PY	Plethodontid modulating factor	113913185	<i>Plethodon yonahlossee</i>	66
CD59.H	CD59 glycoprotein	116021	<i>Homo sapiens</i>	77
CD58.M	CD59 glycoprotein	13878360	<i>Mus musculus</i>	73
CD59.R	CD59 glycoprotein	2507508	<i>Rattus norvegicus</i>	79
Ly6H.H	Lymphocyte antigen 6H	10720070	<i>Homo sapiens</i>	90
Ly6H.CM	Lymphocyte antigen 6H	167008973	<i>Macaca fascicularis</i>	90
Ly6H.B	Lymphocyte antigen 6H	167008972	<i>Bos taurus</i>	90
Ly6H.M	Lymphocyte antigen 6H	10720078	<i>Mus musculus</i>	85
SLURP1.H	Secreted Ly-6/uPAR-related protein 1	3287957	<i>Homo sapiens</i>	81
SLURP1.M	Secreted Ly-6/uPAR-related protein 1	14916717	<i>Mus musculus</i>	88
SLURP2.H	Secreted Ly6/uPAR related protein 2	74727391	<i>Homo sapiens</i>	75
SLURP2.M	Secreted Ly6/uPAR related protein 2	123778205	<i>Mus musculus</i>	75
Lynx1.H	Ly-6/neurotoxin-like protein 1	47117907	<i>Homo sapiens</i>	73
Lynx1.C	Ly-6/neurotoxin-like protein 1	61214436	<i>Pan troglodytes</i>	73
Lynx1.RM	Ly-6/neurotoxin-like protein 1	46576878	<i>Macaca mulatta</i>	73
Lynx1.BM	Ly-6/neurotoxin-like protein 1	75040497	<i>Saimiri boliviensis</i>	73
Lynx1.B	Ly-6/neurotoxin-like protein 1	126256577	<i>Bos taurus</i>	73
Lynx1.M	Ly-6/neurotoxin-like protein 1	24212024	<i>Mus musculus</i>	73

Table 2: Family specific conserved pattern in protein sequences

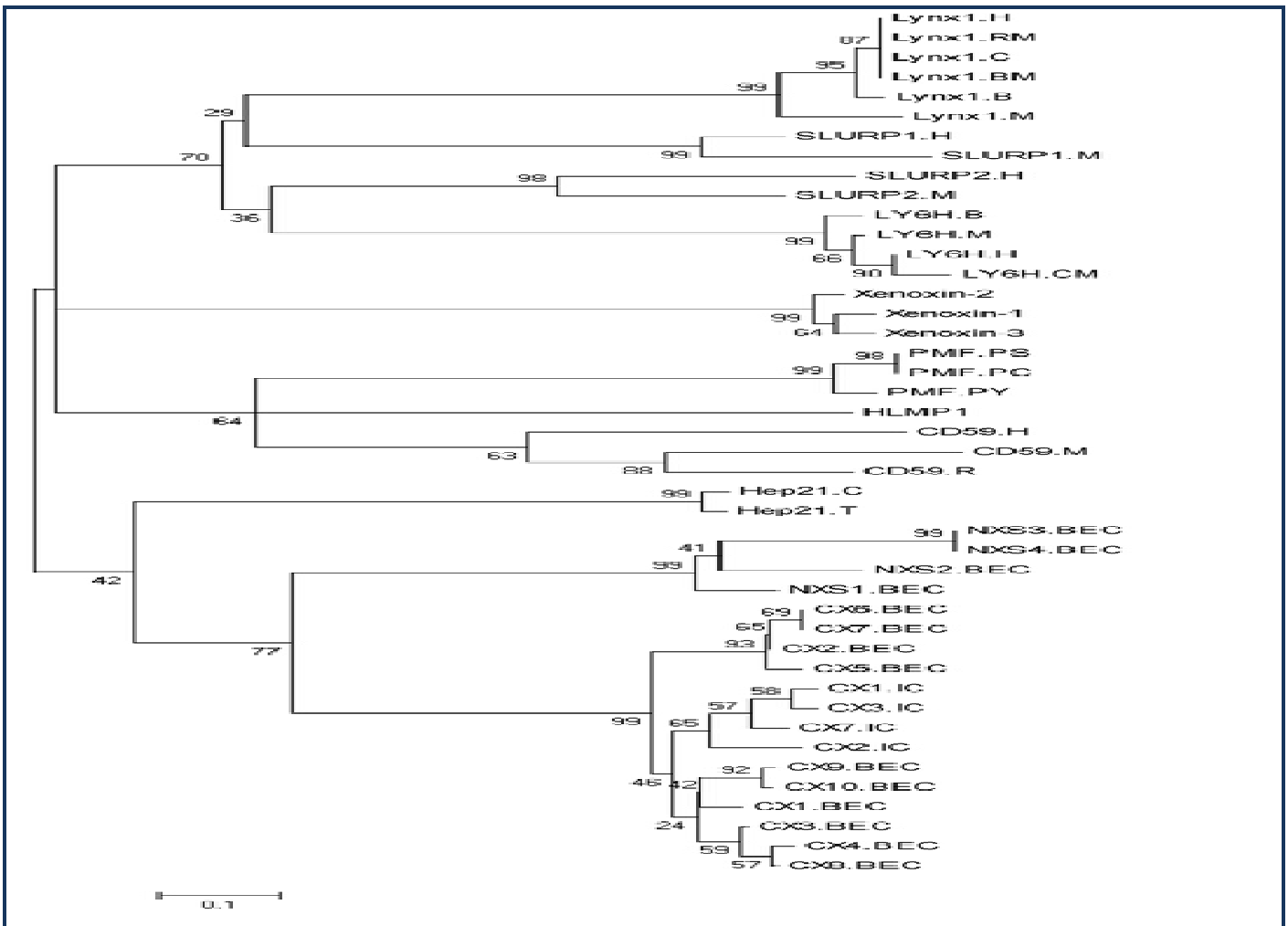
Protein name	Family specific conserved patterns
Cytotoxins of <i>Naja annulifera</i>	K-x-C-P-[EK]-G-K-N-L-C-Y-K-M-[FY]-M-V-[AS]-T-x(2)-[ILV]-P-V-K-R-G-C-I-[DN]-V-C-P-K-[DN]-S-A-L-V-K-Y-[MV]-C-C-[NS]-T-[DN]-[KR]-C-N
Cytotoxins of <i>Naja naja</i>	K-x-C-P-A-G-K-[DN]-L-C-Y-K-M-[FY]-M-V-[AS]-[NT]-x(2)-V-P-V-K-R-G-C-I-D-V-C-P-K-[NS]-S-L-[LV]-[LV]-K-Y-x-C-C-N-T-D-R-C-N
Short neurotoxins	P-[IP]-T-x(0,2)-T-x(0,1)-C-P-G-E-x-N-C-Y-K-K-[QR]-W-[RS]-[DG]-H-R-G-[ST]-I-x-E-R-G-C-G-C-P-S-V-K-K-G-[IV]-[EG]-I-x-C-C-x-T-[DN]-K-C-N
Xenoxin	V-N-L-Q-A-N-G-[IV]-K-M-T-Q-E-C-A-K-E-D-[NT]-K-C-L-T-L-R-S-L-K-K-T-L-K-F-C-A-S-[DG]-R-x-C-x(0,1)-T-x(0,1)-M-K-I-x-S-L-P-G
Hep21	A-I-I-K-T-S-L-G-K-[MV]-T-L-Y-Y-Q-Q-G-C-T-S-A-L-N-C-G-R-E-R-A-S-D-A-E-S-R-L-T-S-R-Y-S-C-C-E-T-D-L-C-N-E
PMF	L-D-G-G-T-E-E-C-L-P-G-N-D-A-V-C-M-Y-Y-K-[FY]-E-x(1,2)-E-x(0,1)-Y-K-D-C-G-G-P-D-A-C-[DT]-E-M-[KR]-D-D-F-T-E-N-E-x-E-H-G
CD59	L-x-C-Y-x-C-x-[DNO]-P-x(3,4)-C-x(2)-[AN]-[SV]-[NT]-C-S-[PS]-[DN]-x-D-[AS]-C-L-x-[AT]-x-[AS]-G-x-Q-V-Y-[NO]-x-C-W-[KR]-x-[ES]-[DH]-C
Ly6H	S-S-R-K-D-H-S-V-N-K-M-C-A-S-S-C-D-F-V-K-R-H-F-F-S-D-Y-L-M-G-F-I-N-S-G-I-L-K-V-D-V-D-C-x-E-K-D-L-C-N
SLURP1	T-[TV]-L-x-T-V-E-A-[AE]-[FY]-P-F-N-x-S-P-x(0,1)-V-x(0,1)-T-R-S-C-S-S-C-[LV]-A-T-D-P-D-[GS]-I-G-[AV]-A-H-[LP]-[IV]-F-C-C-F-R-D-L-C-N-S
SLURP2	H-C-V-x(2)-A-T-R-x(2,3)-S-x-T-x(0,1)-D-L-P-L-V-T-K-M-C-[HY]-x-G-C-P-D-[IV]-[PS]-S-L-G-L-G-P-[HY]-V-S-I-A-C-C-Q-[ST]-[NS]-L-C-N
Lynx1	C-x(4)-Y-N-G-[DE]-N-C-F-x-P-M-R-C-P-A-M-[AV]-[AST]-Y-C-M-T-T-R-T-Y-[FY]-T-P-x-R-M-K-V-[RS]-K-S-C-V-P-[RS]-C-F-E-T



Supplementary figure 1: Multiple sequence alignment of various snake venom toxin proteins and related non-toxin proteins of other chordates (Clustal-X coloring scheme).



Supplementary figure 2: Sequence logo representation of snake venom cytotoxins and short neurotoxins. Sequence logo was created by WebLogo program (<http://weblogo.berkeley.edu/>).



Supplementary figure 3: Phylogram of various snake venom toxin proteins and related non-toxin proteins of other chordates.