Peptides 92 (2017) 31–37

Contents lists available at ScienceDirect

Peptides

journal homepage: www.elsevier.com/locate/peptides

Adaptive evolution of insect selective excitatory β-type sodium channel neurotoxins from scorpion venom

Wenlan Wu[a,*], Zhongjie Li[a], Yibao Ma[b,*]

[a] Medical School, Henan University of Science and Technology, Luoyang, Henan Province, PR China
[b] Department of Biochemistry and Molecular Biology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

ARTICLE INFO

Keywords:
Scorpion neurotoxins
Insect selective excitatory β-NaScTxs
Molecular phylogeny
Adaptive evolution

ABSTRACT

Insect selective excitatory β-type sodium channel neurotoxins from scorpion venom (β-NaScTxs) are composed of about 70–76 amino acid residues and share a common scaffold stabilized by four unique disulfide bonds. The phylogenetic analysis of these toxins was hindered by limited sequence data. In our recent study, two new insect selective excitatory β-NaScTxs, LmIT and ImIT, were isolated from *Lychas mucronatus* and *Isometrus maculatus*, respectively. With the sequences previously reported, we examined the adaptive molecular evolution of insect selective excitatory β-NaScTxs by estimating the nonsynonymous-to-synonymous rate ratio ($\omega = d_N/d_S$). The results revealed 12 positively selected sites in the genes of insect selective excitatory β-NaScTxs. Moreover, these positively selected sites match well with the sites important for interacting with sodium channels, as demonstrated in previous mutagenesis study. These results reveal that adaptive evolution after gene duplication is one of the most important genetic mechanisms of scorpion neurotoxin diversification.

1. Introduction

With millions years of intriguing evolutionary innovation, the venom peptides favored ancient lineages, such as snake, cone snails, centipedes, spiders and scorpions, by powerful mixtures of neurotoxins for predation and defense [1–6]. The scorpion neurotoxins (ScTxs) are extensively investigated for their prominent sequence and function diversification [7–12] in species even separated by genetic and geographic distance [13–17]. Voltage-gated sodium (Na_v) channels are the primary target for ScTxs due to their important role in neuronal excitability in the natural prey, insects [18,19]. Based on their pharmacological and electrophysiological profile, the Na_v channel ScTxs (NaScTxs) can be classified into 2 major groups, α-NaScTxs and β-NaScTxs, binding to receptor site 3 and site 4 of the Na_v channel, respectively [20–25]. Moreover, the β-NaScTxs can be classified to three groups, insect selective β-NaScTxs, mammalian selective β-NaScTxs and β-like NaScTxs [26,27]. The insect selective β-NaScTxs can be further subdivided into excitatory β-NaScTxs and depressant β-NaScTxs based on gene organization and the symptoms in blow fly larvae after β-NaScTxs injection [28]. The excitatory β-NaScTxs, which have been found strictly in "Old World" scorpions, display high selectivity and low-capacity to insect nervous Na_v channels, highlighting them as the potential insecticides [29–31]. Injection of the excitatory β-NaScTxs into fly larvae increased transmitter release and repetitive action potentials in motor nerves, leading to fast contraction paralysis [24,32].

The excitatory β-NaScTxs are composed of about 70–76 amino acid residues and share a common scaffold (cysteine-stabilized α-helix/β-sheet, CSαβ) with four disulfide bonds like other NaScTxs [33,34]. Due to the common scaffold and gene organization, it was suggested that they had radiated from a common ancestor [35]. However, compared with other NaScTxs, the excitatory β-NaScTxs have an atypical disulfide bond pairing between the cysteine residue of the β2 strand (Cys5) and the cysteine residue at the C-terminal (Cys8) [33]. Moreover, several important differences were observed, especially in the C-terminal region which may influence the function of neurotoxin [36,37].

Multigene families of ScTxs have been identified to undergo dynamic molecular duplications followed by positive selection to obtain new functions [38,39]. Previous studies on toxin structure and function provided insights into understanding of the function-driven positive selection of scorpion α-NaScTxs and depressant β-NaScTxs [40,41]. However the influence of positive selection on excitatory β-NaScTxs diversification hasn't been studied yet. In our study, we hypothesized that function-driven positive selection is important for diversifying excitatory β-NaScTxs from different species. Hence, we identified 10 peptide-coding sequences for our analysis and documented the diversity of these genes for the first time.

In the present study, we characterized two new insect selective

* Corresponding authors.
E-mail addresses: whenas@haust.edu.cn (W. Wu), mayibao@163.com (Y. Ma).

http://dx.doi.org/10.1016/j.peptides.2017.03.007
Received 19 December 2016; Received in revised form 25 March 2017; Accepted 27 March 2017
Available online 29 March 2017

excitatory β-NaScTx transcripts from the venom of *Lychas mucronatus* and *Isometrus maculatus*, and used molecular evolution method [42–44] for identifying positively selected sites of the excitatory β-NaScTxs by examining nucleotide sequences from a diversity of species. By PAML program [45], three pairs of models, M2a/M1a, M3/M0 and M8/M7, were compared while the M2a, M3 and M8 assume that some sites are under positive selection. Furthermore, the M7 and M8 models assume a β distribution of the χ classes. The likelihoods of these models were compared using a likelihood ratio test (LRT) with a χ^2 distribution [46]. The Bayes Empirical Bayes approach (BEB) [47] was used to calculate the posterior probabilities (PP) for site classes. A site is considered under positive selection if the PP > 0.95. The match of the evolutionary sites with the functional areas indicates that these sites are involved in recognition and modulation of Na_v channels. The determination can help us to predict the important functional sites and to make rational molecular design [48–50].

2. Materials and methods

2.1. cDNA library construction and sequencing of two new excitatory β-NaScTxs

Two cDNA libraries were constructed from the venom glands of the scorpion species (*Lychas mucronatus* and *Isometrus maculatus*) collected from Hainan, China using previously reported methods [8,49,51,52]. The SuperScript™ Plasmid System (Invitrogen, Carlsbad, CA, USA) was used to construct directional cDNA libraries according to the supplier's instructions. And cDNAs were primed by an oligo(dT)-containing Not *I* primer-adaptor, ligated to a Sal *I* adaptor, and then cloned into the pSPORT-1 plasmid. Multiple transformations yielded libraries comprising approximately 2.5×10^6 colonies. Random colonies from both cDNA libraries were selected and sequenced on ABI3730 automated sequencers (Applied Biosystems, Foster City, CA, USA). The non-redundant sequences were deposited in the dbEST (*Lychas mucronatus*: FE193398-FE193725; *Isometrus maculatus*: FD660398-FD660725).

2.2. Peptide sequence alignment and tree construction

Sixteen amino acid sequences of insect selective excitatory β-NaScTxs, referred to by their GenBank accession numbers, were used including 2 obtained from our recent study and 14 previously published toxins from other seven species (summarized in Table 1). The sequence alignment was performed using MUSCLE online (http://www.ebi.ac.uk/Tools/msa/muscle/) [53]. Phylogenetic trees were constructed using the Neighbor-Joining method using MEGA7 software [54].

Table 1
Sequence information of insect selective excitatory β-NaScTxs.

Peptide names	Scorpion species	Accession Numbers	
		Protein No.	DNA No.
Bj-xtrIT	*Hottentotta judaicus*	P56637	AJ012312
BmKIT1	*Mesobuthus martensii*	O61668	AF057555
BmKIT-AP	*Mesobuthus martensii*	O77091	AF055672
LqhIT1a	*Leiurus quinquestriatus hebraeus*	P68721	[35]
LqhIT1b	*Leiurus quinquestriatus hebraeus*	P68722	
LqhIT1c	*Leiurus quinquestriatus hebraeus*	P68723	
LqhIT1d	*Leiurus quinquestriatus hebraeus*	P68724	
LqqIT1	*Leiurus quinquestriatus quinquestriatus*	P19856	
BoiTx696	*Buthus occitanus israelis*	ACJ23114	FJ360794
BoiTx671.2	*Buthus occitanus israelis*	ACJ23113	FJ360793
Isom1	*Isometrus vittatus*	P0C5H1	
Isom2	*Isometrus vittatus*	P0C5H2	
AaHIT1	*Androctonus australis*	P01497	M27706.1
AaHIT2	*Androctonus australis*	P15147	M27707.1
LmIT	*Lychas mucronatus*	ABX76776	FE193710
ImIT	*Isometrus maculatus*	ACD11769	FD660411

2.3. cDNA sequence alignment and tree construction

To construct trees used for the estimation of ω by the maximum likelihood (ML) method, peptide-coding nucleotide sequences of mature excitatory β-NaScTxs were used and aligned corresponding to their peptide sequences using MUSCLE. The DNA sequence of LqhIT1a was obtained from the reference [35]. The appropriate model of sequence evolution was selected using PhyML with automatic model selection by Smart Model Selection (SMS) to determine the evolutionary model which best fits the input data [55]. Trees were generated based on the Bayesian information criterion (BIC) [56]. Clade stability was estimated by SH-like aLRT branch supports [57] with PhyML. Trees were visualized in Figtree (http://tree.bio.ed.ac.uk/software/figtree/) for further analysis.

2.4. Evolutionary analyses for positive selection

To test the role of positive selection in the evolution of the excitatory β-NaScTxs nucleotide sequences, we performed a ML approaches implemented in the program CodeML of the PAML package (http://abacus.gene.ucl.ac.uk/software/paml.html) to estimate the ω and compare different site-models [43–45]. In all analyses, the tree topology obtained above was used, and likelihood values were estimated. To obtain a reliable result, multiple models were used for detecting positive selection that affects individual sites. Likelihood values of three pairs of models were compared with different assumed ω distributions: M2a/M1a, M3/M0 and M8/M7. Model M0 is the one-ratio model that assumes a single ω ratio for all codon sites. Model M3 uses a discrete distribution of five site classes with the ω ratios estimated from the data and the proportions. The M1a (NearlyNeutral) and M7 (β) models allow the ω ratios of the site to take values in the interval (0, 1). The M2a (PositiveSelection) and M8 (β and ω) add an extra class of sites to the models M7 and M1a, respectively, and assume the sites with ω > 1. Moreover, the M7 and M8 models assume a β distribution of the χ classes. The likelihoods of these six models were compared using LRT with a χ2 distribution [46]. Only when the alternate model shows a better fit than the null model in the LRT, are its results could be considered significant. The BEB was used to detect amino acids under positive selection by PP for site classes. Belonging to the ω > 1 class, the sites with PP > 0.95 are considered positively selected [17]. The results corresponding to the higher likelihood value are used. Sites under positive selection were mapped onto the structures using the UCSF Chimera program (http://www.cgl.ucsf.edu/chimera/).

3. Results

3.1. Molecular characterization of two excitatory β-NaScTxs from scorpions Lychas mucronatus and Isometrus maculatus

After construction of the venomous gland cDNA libraries of *Lychas mucronatus* and *Isometrus maculatus*, sequence analysis showed that 22 clones belonged to β-NaScTxs and two new excitatory β-NaScTxs FE193710 and FD660411 of the scorpions were isolated from the libraries by previous methods [58–60]. Two of the obtained nucleotide sequences displayed ORFs encoding new putative neurotoxins which were termed LmIT and ImIT. The precursor nucleotide sequences have three parts: 5′ UTR, ORF and 3′ UTR. The 5′ and 3′ UTRs of LmIT have 6 and 90 bp (Fig. 1A), respectively. The 5′ and 3′ UTRs of ImIT have 27 and 74 bp (Fig. 1B), respectively. At the 3′ UTR end of the cDNAs, the AATAAA polyadenylation signals of both LmIT and ImIT are found 18 nt upstream of the poly(A) tails. LmIT contains an ORF of 267 bp encoding a precursor of 89 residues (Fig. 1A) and ImIT has an ORF of 270 bp for 90 residues (Fig. 1B). SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) predicted that the precursor of LmIT and ImIT contain putative signal peptides of 18 and 19 residues, respectively, followed by mature β-NaScTxs of 71 residues.