FEBS Letters xxx (2014) xxx–xxx

journal homepage: www.FEBSLetters.org

Genes and evolution of two-domain toxins from lynx spider venom ☆

Maria Y. Sachkova [a,*], Anna A. Slavokhotova [a,b], Eugene V. Grishin [a], Alexander A. Vassilevski [a,*]

[a] M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya, 117997 Moscow, Russian Federation
[b] N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 3 Gubkina, 119991 Moscow, Russian Federation

ARTICLE INFO

Article history:
Received 30 October 2013
Revised 2 January 2014
Accepted 10 January 2014
Available online xxxx

Edited by Takashi Gojobori

Keywords:
Molecular evolution
Spider venom gland
Inhibitor cystine knot (ICK)
cDNA library
Toxin gene

ABSTRACT

Spiderines are comparatively long polypeptide toxins (~110 residues) from lynx spiders (genus *Oxyopes*). They are built of an N-terminal linear cationic domain (~40 residues) and a C-terminal knottin domain (~60 residues). The linear domain empowers spiderines with strong cytolytic activity. In the present work we report 16 novel spiderine sequences from *Oxyopes takobius* and *Oxyopes lineatus* classified into two subfamilies. Strikingly, negative selection acts on both linear and knottin domains. Genes encoding *Oxyopes* two-domain toxins were sequenced and found to be intronless. We further discuss a possible scenario of lynx spider modular toxin evolution.

1. Introduction

Spider venom is a source of biologically active polypeptides composing combinatorial libraries [1,2]. Toxin variability is not limited to multiplicity of amino acid sequences within the confines of the same structural motif. In addition to the more common one-domain peptide toxins, elaborate two-domain (or "modular") toxins have been recognized recently.

One-domain toxins usually possess the inhibitory cystine knot (ICK, or knottin) fold, or they are just short linear molecules adopting amphipathic α-helical conformation, with neurotoxic or cytolytic activities, correspondingly. Such "common" toxins may be used as building blocks to construct more complex two-domain molecules. Toxins corresponding to all possible combinations of these structural elements have been described in the venoms of spiders from different families: ICK-ICK (DkTx [3] and CpTx [4]), linear-linear (cyto-insectotoxins [5]), ICK-linear (LtTx, or latartoxins [6], and CsTx [7]), and linear-ICK (spiderines [8]). Intriguingly, different scorpions were also found to produce two-domain toxins referred to as scorpines and β-KTxs [9,10].

Spiderines (OtTx) from the lynx spider *Oxyopes takobius* are long polypeptides (~110 residues) built of an N-terminal linear domain (~40 residues) and a C-terminal ICK domain (~60 residues) linked by a short sequence resembling a mutated processing site [8]. OtTx possess strong cytolytic activity due to the linear domain. The functional role of the knottin domain is yet to be established. The "chimeric" structure of spiderines raises questions about their evolutionary origin.

Oxyopes lineatus is a species closely related to *O. takobius*. Interestingly, both spiders produce the same major single-domain ICK neurotoxin oxytoxin-1 (OxyTx 1) [11,12]. However, a variety of linear cytotoxins (oxyopinins) are present in *O. takobius* venom [11,13] but not in *O. lineatus* that conversely produces oxytoxin-2 (OxyTx 2) absent in *O. takobius* [12]. We therefore decided to mine *O. lineatus* venom glands for spiderines.

Genes encoding spider toxins are poorly studied in part due to the lack of data on spider genome sequences. In the present work we report sequences of genes encoding lynx spider two-domain toxins. Based on these data, we attempt to elucidate a possible scenario of *Oxyopes* two-domain toxin evolution.

2. Materials and methods

2.1. DNA and RNA purification from O. lineatus venom glands

Specimens of *O. lineatus* spider were collected in Kazakhstan by Andrey Feodorov (Fauna Laboratories, Ltd., Republic of Kazakhstan). Venom glands were excised and stored at −70 °C. ~100 mg

Abbreviations: gDNA, genomic DNA; ICK, inhibitor cystine knot; LRT, likelihood ratio test; ML, maximum likelihood; NG method, Nei–Gojobori method; PQM, processing quadruplet; RACE, rapid amplification of cDNA ends

☆ Nucleotide sequences reported in this paper have been deposited in GenBank with accession numbers KF766543–KF766563.

* Corresponding authors. Fax: +7 495 330 73 01.
E-mail addresses: sachkovamasha@mail.ru (M.Y. Sachkova), avas@ibch.ru (A.A. Vassilevski).

http://dx.doi.org/10.1016/j.febslet.2014.01.018
0014-5793/$36.00 © 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

of venom glands from several individuals were used to extract genomic DNA (gDNA) and total RNA. The gland tissue was dissolved at +55 °C in 500 μl of 50 mM Tris–HCl (pH 8.0), 200 mM NaCl, 100 mM EDTA, and 1% SDS solution containing 1 mg/ml proteinase K. 250 μl of 5 M NaCl were added and the pellet was removed. DNA/RNA mixture precipitation by ethanol was followed by phenol–chloroform extraction. After repeated ethanol precipitation the DNA/RNA mixture was dissolved in DNAase/RNAase-free water. The obtained gDNA was consequently used in PCR reactions. To obtain cDNA, mRNA was reverse transcribed by the MINT kit (Evrogen, Russia) following the manufacturer's recommendations. All cDNA sequences were flanked by adaptor sequences that were then used for rapid amplification of cDNA ends (RACE) (see below).

2.2. PCR amplification of cDNA and gDNA fragments and sequencing

Based on cDNA sequences from the *O. takobius* cDNA library, specific primers were designed to amplify two-domain toxin genes from *O. lineatus* (Table 1). 5′ and 3′ RACE were carried out with the universal primer T7cap and a corresponding specific primer. PCR products were cloned into the pAL-TA plasmid vector (Evrogen) and sequenced (see below). Using the newly established cDNA sequences we designed primers to amplify gDNA fragments encoding *O. lineatus* two-domain toxins (Table 1).

PCR reactions were performed with the Evrogen PCR kit. Purified PCR products were ligated into the pAL-TA plasmid vector that was used for transformation of competent *Escherichia coli* XL1-Blue cells. Positive clones were sequenced with the M13 forward primer. The Lasergene package (DNASTAR, USA) was used for analysis of DNA sequencing results and other manipulations with nucleotide sequences.

2.3. Analysis of cDNA sequences

O. takobius venom gland cDNA library was constructed in collaboration with DuPont Agriculture and Nutrition [8]. *O. lineatus* cDNA was synthesized as described above. To translate cDNA sequences *in silico*, the DNASTAR software was used. Signal peptides were identified by the SignalP 4.1 online tool (http://www.cbs.dtu.dk/services/SignalP/). Propeptide sequences were assigned as preceding the processing quadruplet motif (PQM) cleavage sites [14]. Multiple and pairwise alignments were constructed by the ClustalW program [15] using Vector NTI Suite 8 (Life Technologies, USA) and MEGA 5 software [16]. Search for putative splice sites was performed by the Fruitfly Splice Predictor (http://www.fruitfly.org/seq_tools/splice.html) [17].

The following analyses were performed using MEGA 5. For phylogenetic trees, the evolutionary history was inferred using the neighbor-joining method [18]. The bootstrap test (500 replicates) was performed to calculate the percentage of replicate trees in which the associated sequences clustered together. The evolutionary distances were computed using the maximum composite likelihood method [19]. Codon-based *Z*-test of Purifying/Positive Selection was performed using the Nei–Gojobori (NG) method [20].

Further phylogenetic analysis was executed with the maximum likelihood (ML) method by the CODEML program from the package PAML (version 4.7a) [21] using the graphical interface PAML X [22]. All alignment gaps were deleted for the analyses. For pairwise comparisons, the program was run with runmode = −2 and model M0. For likelihood ratio tests (LRTs), the program was run with runmode = 0 and five models: M0, M1a, M2a, M7, and M8. Branch lengths were estimated by the model M0 and used in all subsequent analyses. The natural logarithm of the maximum likelihood (lnL) was calculated for each model. LRTs were constructed for the pairs M1a/M2a and M7/M8. For each LRT, $2\Delta\ell = 2 \times (\ln L_1 - \ln L_0)$ was computed, where L_0 is the model that does not allow for positive selection (M1a and M7). $2\Delta\ell$ values were compared against χ^2 table for 2 degrees of freedom (critical χ^2 for 2 degrees of freedom and $P < 0.05$ or $P < 0.1$ is 5.9915 or 4.6052, correspondingly).

Similarity search was performed by the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) in the Nucleotide and Non-redundant protein sequences NCBI databases. Nucleotide and amino acid sequences of separate domains were used as queries.

3. Results

3.1. Two-domain toxins in lynx spiders

We have recently reported four cDNA sequences of *O. takobius* spiderines [8]. However, the variability of two-domain toxins is wider: in this work we found 9 new sequences belonging to the spiderine family in the *O. takobius* venom glands cDNA library and 6 novel cDNA sequences encoding *O. lineatus* two-domain toxins (OlTx) (GenBank Nos. KF766543–KF766559, KF766561) (see Fig. 1 for the phylogenetic tree).

All the cDNA sequences were translated into protein precursor sequences and processed *in silico* (an example for OlTx 2a is shown in Fig. 2). In total, 15 novel *Oxyopes* two-domain toxin sequences were established belonging to the spiderine family. As many toxins described earlier [1,23], spiderines possess different variation levels in signal, pro- and mature peptides. Signal peptides and propeptides are identical to each other or show a high similarity level (>74% and >62% of identical residues, correspondingly), while mature toxins are more variable (>42% identity). Mature polypeptides can be divided into two subfamilies differing in length: Ox-I (109–114 residues) and Ox-II (135–142 residues). Apart from OtTx 1a, 1b, 2a and 2b reported earlier, 10 newly established sequences were placed to the Ox-I subfamily: OtTx 1c–e, OtTx 2c–f, and OlTx 1a–c. OtTx 3a and 3b and OlTx 2a–c were placed to the Ox-II subfamily. Inside each subfamily, identity between toxins exceeds 68%, while between the subfamilies it is in the range of 42–50%. All Ox-I polypeptides (including OtTx reported earlier [8]) resemble complex precursors with mutated processing sites, but none of Ox-II polypeptides contains sequences of high similarity to the PQM processing site.

Length difference between the Ox-I and Ox-II subfamilies is due to the N-terminal domains of the mature polypeptides: they are composed of 47–55 and 69–73 residues, correspondingly (Fig. 3).

Table 1
Oligonucleotide primers used for amplification of cDNA and gDNA fragments encoding OlTx toxins.

Gene	Template	Direction	Sequence
OlTx1	gDNA	Forward	CGATTTCAAAATGAAGTTCTCTTTGG
		Reverse	TTGAGCAAGAATTCTATACAACTG
	cDNA	Forward	AAGGGTCTAGAGAAAGCAACACC
		Reverse	ATACAACTGGGAGGTCACAGGTG
OlTx2	gDNA, cDNA	Forward	AACATCGTTTTCAAAATGAAGATCG
		Reverse	AGTCTACTCAGCCGCAGGTTCCT
Both	cDNA	Universal primer T7cap	GTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT

Please cite this article in press as: Sachkova, M.Y., et al. Genes and evolution of two-domain toxins from lynx spider venom. FEBS Lett. (2014), http://dx.doi.org/10.1016/j.febslet.2014.01.018