Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimie 88 (2006) 131–140
BIOCHIMIE
www.elsevier.com/locate/biochi

Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins

Canhui Pi, Yun Liu, Can Peng, Xiuhua Jiang, Junliang Liu, Bin Xu, Xuesong Yu, Yanghong Yu, Xiaoyu Jiang, Lei Wang, Meiling Dong, Shangwu Chen, An-Long Xu *

State Key Laboratory of Biocontrol, Guangdong Province Key Laboratory of Therapeutic Functional Genes, The Open Laboratory for Marine Functional Genomics of State High-Tech Development Program, Department of Biochemistry, College of Life Sciences, Sun Yat-sen (Zhongshan) University, 135 Xingangxi Road, 510275 Guangzhou, China

Received 21 May 2005; accepted 16 August 2005

Available online 08 September 2005

Abstract

Cone snails (genus *Conus*) are predatory marine gastropods that use venom peptides for interacting with prey, predators and competitors. A majority of these peptides, generally known as conotoxins demonstrate striking selectivity in targeting specific subtypes of ion channels and neurotransmitter receptors. So they are not only useful tools in neuroscience to characterize receptors and receptor subtypes, but offer great potential in new drug research and development as well. Here, a cDNA library from the venom ducts of a fish-hunting cone snail species, *Conus striatus* is described for the generation of expressed sequence tags (ESTs). A total of 429 ESTs were grouped into 137 clusters or singletons. Among these sequences, 221 were toxin sequences, accounting for 52.1% (corresponding to 19 clusters) of all transcripts. A-superfamily (132 ESTs) and O-superfamily conotoxins (80 ESTs) constitute the predominant toxin components. Some non-disulfide-rich *Conus* peptides were also found. The expression profile of conotoxins also explained to some extent the pharmacological and physiological reactions elicited by this typical piscivorous species. For the first time, a nonstop transcript of conotoxin was identified, which is suggestive that alternative polyadenylation may be a means of post-transcriptional regulation of conotoxin production. A comparison analysis of these conotoxins reveals the different variation and divergence patterns in these two superfamilies. Our investigations indicate that focal hypermutation, block substitution and exon shuffling are three main mechanisms leading to the conotoxin diversity in a species. The comprehensive set of *Conus* gene sequences allowed the identification of the representative classes of conotoxins and related components, which may lay the foundation for further research and development of conotoxins.

Keywords: Cone snail; Conopeptide; Expressed sequence tag; Toxin cable; Nonstop transcript; Evolution mechanism; Nomenclature

1. Introduction

Cone snails (genus *Conus*) are predatory gastropods that are found in tropic marine habits around the world. They are well known for an unusually large number of small, highly structured peptides (conopeptides or/and conotoxins) that are secreted in the venom duct. These venom peptides were originally developed as part of an envenomation survival strategy for prey capture, defense and competitive interactions [1]. During the history of evolution, they have been optimized to target specific ion channels and neurotransmitter receptors with extraordinary high affinity and selectivity, especially discriminating between closely related receptor subtypes [2,3]. As a consequence, conotoxins are not only useful pharmacological tools to characterize receptors and receptor subtypes in neuroscience research, but offer great potential in new drug research and development as well.

From the evolutionary point of view, cone snails are among the most successful marine animals, comprising arguably the

Abbreviations: AChR, acetylcholine receptors; BLAST, basic local alignment search tool; BLAST N, nucleotide–nucleotide BLAST; BLAST P, protein–protein BLAST; bp, base pairs; EST, expressed sequence tag; NCBI, National Center for Biotechnology Information; NSD, nonstop-mediated decay; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; UTR, untranslated regions.

* Corresponding author. Tel.: +86 20 8411 3655; fax: +86 20 8403 8377.
E-mail address: ls36@zsu.edu.cn (A.-L. Xu).

0300-9084/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.biochi.2005.08.001

largest single genus of venomous animals known. At present, it is a rough estimate that there are more than 500 *Conus* species, each expressing a repertoire of 50–200 conopeptides, with little overlap between species. Although these *Conus* peptides comprise an unprecedented vast number of unique peptide sequences, the majority of all conopeptides appear to be derived from a dozen gene superfamilies, and members of each superfamily share a highly conserved signal sequence in their precursors and characteristic pattern of disulfide pairing [1,4]. How *Conus* peptide hypervariability may have occurred in such a short evolutionary history has been pursued for a long time, but until now, relatively little is known.

Analysis of expressed sequence tags (ESTs) is a powerful approach for novel gene identification, homologous gene comparison, transcription profiling and some other applications. There are mainly two strategies in EST analysis. Some studies generated a sufficient number of cDNA clones and sequenced thousands of ESTs to characterize even rare ones [5]. Others conduct the sequencing of hundreds to a thousand transcripts to create a partial but representative catalogue of expressed sequences. The latter approach appears to be adequate and efficient for the identification of some specified tissues such as snake venom glands [6,7] and porcine skeletal muscle [8], since the special proteins are the dominant components and there are so many repeated sequences in the specific cDNA library. Apparently in our research, sequencing hundreds of ESTs to profile gene expression in cone snail is adequate.

C. striatus is a widespread fish-hunting cone snail. So far, four A-superfamily conotoxins (SI, SIA, SII and SIVA) along with three O-superfamily toxins (SVIA, SVIB and SVIE) and two non-disulfide-rich conopeptides (conopressin-S, arg-conopressin S) have been isolated and characterized [9–16]. The coding sequences of some presumed new toxins were also reported recently [17–19]. However, these studies were mostly focused on the isolation and identification of single peptides through methods of classical biochemistry and cloning of the corresponding cDNA by PCR. The entire spectrum of conopeptides contained in this species has not been identified. Furthermore, the molecular properties of the toxins have not been elucidated systematically.

In an effort to unveil the toxicological and pharmacological potential of conotoxins, and to discuss the probable mechanism of *Conus* peptide evolution and diversification at the molecular level, especially in a single species of cone snails, we constructed a cDNA library from the venom ducts of a fish-hunting cone snail species, *Conus striatus* for the generation of ESTs. Our partially sequenced ESTs allowed the identification of the most common classes of conotoxins and new ones, which may lay the foundation for further research and development of conotoxins. In addition, this library represents the first comprehensive set of *Conus* gene sequences described so far.

2. Materials and methods

2.1. cDNA library construction

Specimens of *C. striatus* were collected from Yalong Bay, the south end of Hainan Island, China. The venom ducts were dissected from living snails, then immediately frozen in liquid nitrogen. The frozen tissues were homogenized and dissolved in TRIZOL®. The following steps for total RNA isolation and cDNA library construction were performed as previously described [20] except the multiple cloning site sequence of plasmid vector pcDNA3.0 was partially replaced by λTriplEx2 (Clontech) segment through the common restriction endonuclease sites EcoR I/Not I to introduce the Sfi I restriction sites.

cDNA size fractionation was performed with CHROMA SPIN-400 provided with library construction kits and the fractions longer than 400 bp were collected. The presence and size of inserts were tested by PCR of randomly chosen colonies using vector primer (T7 forward and SP6 reverse primer). Further verification was performed by digestion with the restriction enzyme Sfi I of total cloned plasmid and analyzed by electrophoresis. The *C. striatus* venom duct library contained 1.1×10^6 independent clones based on our estimation.

2.2. Expressed sequence tags sequencing

cDNA clones were randomly picked and cultured in individual wells of 96-well plates containing appropriate 2 × YT medium with 100 mg/l ampicillin. After overnight culture, plasmids were prepared by using Vitagene 96-easy plasmid Mini-prep kits (Vitagene Biochemical Technique). Purified plasmid DNAs were single-pass sequenced from the 5′-end in an automated ABI PRISM 3700 sequencer (Perkin Elmer) using the T7 promoter primer and ABI PRISM® Big Dye™ terminator v3.1 ready reaction cycle sequencing kits (Applied Biosystems).

2.3. Sequence data analysis

Prior to further analysis, sequencing outputs were trimmed by removal of vector, primer sequences and poly (A) tails with ABI PRISM® DNA Sequencing Analysis Software V.3.3 [21]. Low quality segments and the inserts less than 100 bp were also subjected to removal. All valid sequences were assembled into clusters with a minimal score of 95 using software. The consensus sequences of each cluster were further filtered by screening for homology to ribosomal RNA, mitochondrial DNA and *E. coli* genome sequences [22–24]. After deleting matches, the remaining sequences were used as a query to search nucleotide and protein database available in NCBI by the BLAST N and BLAST X algorithm, respectively [25]. In the following analysis, annotations of possible protein-coding genes were carried out in detail for further study.