ELSEVIER

Toxicon 49 (2007) 1135–1149

TOXICON

www.elsevier.com/locate/toxicon

From the identification of gene organization of α conotoxins to the cloning of novel toxins

Duo-Duo Yuan[a,1], Yu-Hong Han[a,b,1], Chun-Guang Wang[a,*], Cheng-Wu Chi[a,b,**]

[a]*Institute of Protein Research, TongJi University, 1239 Siping Road, Shanghai 200092, China*
[b]*Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Graduate School of the Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China*

Received 30 September 2006; received in revised form 5 February 2007; accepted 5 February 2007
Available online 1 March 2007

Abstract

In the venoms of cone snails, α conotoxins are competitive antagonists of nicotinic acetylcholine receptors. Eleven novel cDNA and eight partial gene sequences (including two pseudogenes) of α conotoxins were identified from five species of cone snail. As expected, every cDNA encodes a precursor of prepropeptide. In all the partial genes of α conotoxins identified, there is a long intron inserted at a fixed position in the pro-region, dividing the encoding region into two exons. The mutation rate in exon I (encoding the signal peptide and a part of pro-region) is much lower than that in exon II (encoding the other part of pro-region, the mature peptide and 3′ untranslational region). Interestingly, the sequences at the 5′ and 3′ end of introns are highly conserved. In addition, in the identified introns exist long dinucleotide (e.g. "GT", "CA") or trinucleotide ("CAT") repeats. In the special case of Pu1.1, there are five almost identical repeats of a 150 bp sequence in the long intron. Taking advantage of the conserved 3′ end sequence of intron, 16 α conotoxins, as well as a pseudogene and three κA conotoxins, were identified from their genomic DNAs. Based on the comparison of these cDNA and gene sequences, a hypothesis of the α conotoxin evolution was proposed.

Keywords: α conotoxin; cDNA cloning; Gene cloning; Intron; Evolution

1. Introduction

The small cysteine-rich conotoxins are used by cone snails to capture prey or defend predators. These toxins can selectively target receptors or ion channels in neuromuscular system (Terlau and Olivera, 2004). Based on the highly conserved signal peptide, conotoxins are classified into seven major superfamilies (O-, M-, A-, S-, T-, P- and I-superfamily). Furthermore, according to the Cys framework and their different targets, conotoxins in the same superfamily can be further divided into different families. For example, O-superfamily currently consists of four families, namely δ, μO, ω and κ conotoxins. Although they share the same Cys pattern —C-C-CC-C-C-, toxins in the first two families specifically affect the inactivation and activation of Na^+ channels, while the latter two family toxins target Ca^{2+} and K^+ channels, respectively.

*Corresponding author. Tel.: +86 21 65984347; fax: +86 21 65988403.
**Corresponding author. Institute of Protein Research, TongJi University, 1239 Siping Road, Shanghai 200092, China. Tel.: +86 21 54921165; fax: +86 21 54921011.
E-mail addresses: chunguangwang@mail.tongji.edu.cn (C.-G. Wang), chi@sunm.shcnc.ac.cn (C.-W. Chi).
[1]These authors contribute equally to this paper.

0041-0101/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.toxicon.2007.02.011

α conotoxins, belonging to A-superfamily, are competitive antagonists of muscular and neuronal nicotinic acetylcholine receptors (*n*AChRs). They have a conserved -CC-C-C- cysteine framework (Santos et al., 2004) and can be further divided into several subfamilies based on the number of residues between the discontinuous cysteines. α3/5 conotoxins are unexceptionally characterized from piscivorous cone snails and bind the muscular *n*AChRs with high affinity (Gray et al., 1981; Favreau et al., 1999). α4/7 conotoxins distribute in cone snails of all three major feeding types (fish-hunting, mollusk-hunting and worm-hunting), potently block certain subtypes of neuronal and muscular *n*AChRs (Cartier et al., 1996; Loughnan et al., 1998). Several minor groups of α conotoxins have also been reported, such as α4/3 and α4/6 subfamily (McIntosh et al., 1994; Luo et al., 1998). Recently, a novel α4/4 conotoxin BuIA has been identified from a fish-hunting cone snail *Conus bullatus* by gene cloning. (Azam et al., 2005).

Most conotoxins, including α conotoxins, are initially translated as large precursors (70–120 amino acids) which have a characteristic prepropeptide organization. In general, members from the same superfamily share a highly conserved signal peptide (pre-region) and a rather conserved pro-region, however the C-terminal toxin-encoding region is highly variable. It was mentioned that there was a large intron in the pro-region of α conotoxin (Olivera et al., 1997). However, till now the detailed information about the gene sequence of α conotoxin has not been reported yet.

Here we report the identification of eleven novel cDNA sequences (five of them with full length 3′ untranslated region) and six partial gene sequences of α conotoxins plus two pseudogenes from five species of cone snails of all three major feeding types. All these partial gene sequences share the same organization: two exons and one intron in between. Though varying in length, these introns have the same inserting position in the pro-regions. Moreover, 16 α conotoxins and one pseudogene were also characterized from genomic DNAs of five species of cone snails based on the conserved 3′ end sequence of introns.

2. Materials and methods

2.1. Materials

Specimens of *C. marmoreus, C. leopardus, C. achatinus, C. pulicarius* and *C. quecrinus* were collected from South China Sea near Sanya, Hainan Province. The genomic DNAs were extracted from tissues using Genomic-tip 20/G (QIAGEN, Hilden, Germany). *Taq* DNA polymerase was purchased from MBI and pGEM-T Easy vector system from Promega. IPTG (isopropyl-thio-b-D-galactoside), X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactoside) and other reagents were of analytical grade.

2.2. Preparation of total RNAs

Venom ducts frozen in liquid nitrogen were ground into fine powder and homogenized. By using TRIZOL Reagent kit, the total RNA extraction was carried out according to the instruction manual.

2.3. Preparation of genomic DNA

The genomic DNA was extracted following the manufacture's protocol for animal tissue. About 20 mg of muscle tissue frozen in liquid nitrogen was ground into fine powder and homogenized.

2.4. cDNA cloning

About 5 μg total RNA from the venom duct was taken to convert total mRNA into cDNA using Superscript II reverse transcriptase with a universal oligo(dT)-containing adapter primer 5′-GGCCACGCGTCGACTAGTAC$(dT)_{17}$-3′. With the total cDNA as the template, the 5′ forward primer P1 (5′-ATGGGCATGCGGATGATGTT-3′) and 3′ reverse primer P2 (5′-GTCGTGGTTCAGAGGGTC-3′) or P3 (5′-GGCCACGCGTCGACTAGTAC-3′) were used to amplify α conotoxin cDNAs (Fig. 1). P1 and P2 are designed based on the known conserved sequences in the signal peptide and 3′ untranslational region (UTR) of α conotoxin, respectively (Olivera et al., 1996). While P3 is an abridged universal amplification primer devoid of the Poly dT tail. PCR amplification was performed with a cycling protocol of an initial denaturation of 94 °C for 2 min, 10 cycles of 94 °C for 45 s, 55 °C for 45 s, 72 °C for 1 min with the annealing temperature decreasing 1 °C per cycle, then 25 cycles of 94 °C for 45 s, 45 °C for 45 s, 72 °C for 1 min and terminated with a final extension at 72 °C for 7 min.

2.5. Gene cloning

For PCR amplification of the genes of α conotoxins, primers P1 and P2 were used again.

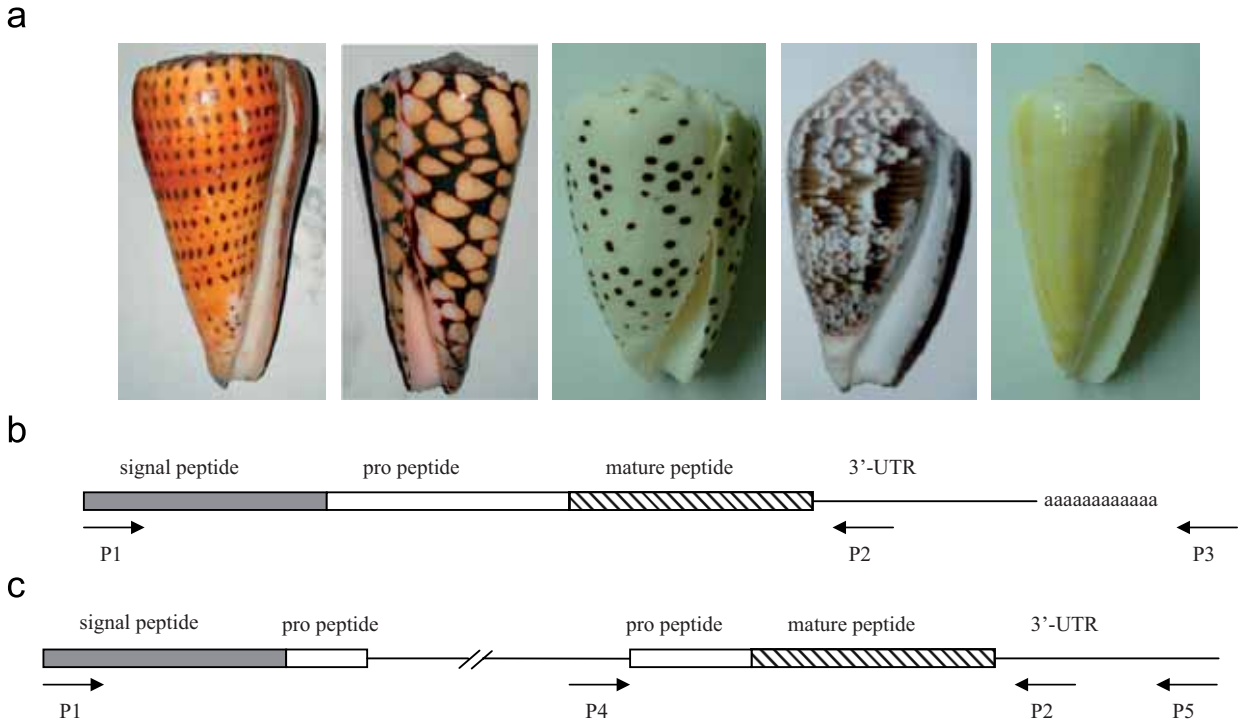

Fig. 1. Cone species studied in this work (a) and diagram of strategy used to clone the sequences of α conotoxin from (b) cDNA, (c) genomic DNA. In panel (a), from left to right are *Conus leopardus, Conus marmoreus, Conus pulicarius, Conus achatinus* and *Conus quercinus*.

Another primer pair used to get more novel sequences of α conotoxins from genomic DNA was P4 (5′-GTGGTTCTGGGTCCAGCA-3′), corresponding to the conserved 3′ end sequence of intron, and P2. The design of primer P5 (5′-TGGACGATGTAATAACAGCAAG-3′) together with 5′ forward primer P1 for Lp1.6a was based on the sequence from 3′UTR of Lp1.1 (Fig. 1). PCR amplification was performed with a cycling protocol of an initial denaturation of 94 °C for 10 min, 30 cycles of 94 °C for 90 s, 50 °C for 60 s, 72 °C for 2 min and terminated with a final extension at 72 °C for 10 min.

2.6. cDNA and genomic DNA sequencing

The PCR products were analyzed by electrophoresis on agarose gel. Then the target band was excised from the gel and purified with Gel Extraction Mini Kit (Beyotime, Haimen, China). The purified PCR products were inserted into the pGEM-T Easy vector via TA cloning for DNA sequencing.

3. Results and discussion

3.1. cDNA cloning of α conotoxins

The conservation of the signal peptide and the UTR of each conotoxin superfamily has been often used to clone novel conotoxins without knowing the sequence of the mature toxins (Lu et al., 1999; Sandall et al., 2003; Wang et al., 2003; Han et al., 2005). With the same strategy, nine cDNAs of novel α conotoxins were identified from *C. marmoreus*, *C. leopardus*, *C. quercinus* and *C. achatinus* with primers P1 and P2 (Fig. 1, Table 1). Six of them (Mr1.1, Mr1.2, Lp1.1, Lp1.4, Qc1.1a and Qc1.1b) belong to α4/7 subfamily. Ac1.1a and Ac1.1b from a piscivorous cone snail *C. achatinus* are the members of α3/5 subfamily. Unexpectedly, we also got a α4/4 conotoxin Qc1.2, from a vermivorous cone snail *C. quercinus*. It is noteworthy that Qc1.2 shares no sequence homology with the other reported α4/4 conotoxin BuIA except the conserved cysteine framework.

All the above cDNAs encode a precursor of prepropeptide, including a highly conserved signal

Table 1
All the predicted sequences of α conotoxins cloned in this study

Conotoxin	Conus species	cDNA cloning		Gene cloning		Predicted mature toxin	Branch	Genebank number
		Primers	3′UTR (bp)	Primers	Introns (bp)			
Ac1.1a	*C. achatinus*	P1,P2				NGR**CC**HPA**C**GKHFN**C**[a]	α3/5	DQ359138
Ac1.1b	*C. achatinus*	P1,P2				NGR**CC**HPA**C**GKHFS**C**[a]	α3/5	DQ359139
				P1,P2	1299			DQ311055
				P4,P2				DQ311072
Ac4.2	*C. achatinus*			P4,P2		APWMVVTATTN**CC**GYTGPA**C**HP**C**L**C**TQS**C**[a]	κA	DQ311073
Ac4.3a	*C. achatinus*			P4,P2		QKELVVTATTT**CC**GYNPMTS**C**PR**C**MCDSS**C** NKKKKP[a]	κA	DQ311074
Ac4.3b	*C. achatinus*			P4,P2		QKELVPSKITT**CC**GYSPGTA**C**PS**C**M**C**TNT**C** KKKNKKP[a]	κA	DQ311075
Mr1.1	*C. marmoreu*	P1,P2		P4,P2		G**CC**SHPA**C**SVNNPDI**C**[a]	α4/7	AY580325
								DQ311077
Mr1.2	*C. marmoreus*	P1,P2		P1,P2	796	G**CC**SNPP**C**YANNQAY**C**N[a]	α4/7	DQ359140
Mr1.3	*C. marmoreus*					G**CC**SHPA**C**RVHYPHV**C**Y[a]	α4/7	DQ309774
Lp1.1	*C. leopardus*	P1,P2	225	P1,P2	791	G**CC**ARAA**C**AGIHQEL**C**[a]	α4/7	DQ359141
		P1,P3						DQ309776
Lp1.4	*C. leopardus*	P1,P2		P1,P2	937	G**CC**SHPA**C**SGNHQEL**C**D[a]	α4/7	AY580324
								DQ311056
Lp1.6a	*C. leopardus*			P1,P5	824	QF**CC**GHYD**C**DFIPNV**C**[a]	α4/6	DQ311057
				P4,P2				DQ359145
Lp1.6b	*C. leopardus*			P4,P2		QI**CC**GYGD**C**GFVPNV**C**V	α4/6	DQ359143
Lp1.7	*C. leopardus*	P1,P3	658			GMWDE**CC**DDPP**C**RQNNMEH**C**PAS	α4/7	DQ311062
Lp1.8	*C. leopardus*	P1,P3	689			GVWDE**CC**KDPQ**C**RQNHMQH**C**PAR	α4/7	DQ311063
Lp1.9	*C. leopardus*			P4,P2		**CC**SDSD**C**NANHPDM**C**S	α4/7	DQ311076
Lp1.10	*C. leopardus*			P4,P2		ND**CC**HNAP**C**RNNHPGI**C**[a]	α4/7	DQ359144
Qc1.1a	*C. quercinus*	P1,P2	820			DE**CC**PDPP**C**KASNPDL**C**DWRS	α4/7	DQ311060
		P1,P3						
Qc1.1b	*C. quercinus*	P1,P2	820	P4,P2		NE**CC**DNPP**C**KSSNPDL**C**DWRS	α4/7	DQ311061
		P1,P3						DQ311064
Qc αL-1	*C. quercinus*			P1,P2	1375	F**C**SDPP**C**RISNPES**C**GWEP	α4/7	DQ311058
				P4,P2				DQ311065
Qc αL-2	*C. quercinus*			P1,P2	1441	GF**C**SDPS**C**RFGNPEL**C**DWRR	α4/7	DQ311059
Qc1.1c	*C. quercinus*			P4,P2		DD**CC**PNPP**C**KASNPDL**C**DWRS	α4/7	DQ311066
Qc1.2	*C. quercinus*	P1,P2		P4,P2		Q**CC**ANPP**C**KHVN**C**[a]	α4/4	AY580320
								DQ311067
Qc1.4a	*C. quercinus*			P4,P2		DG**CC**SNPS**C**SVNNPDI**C**[a]	α4/7	DQ311068
Qc1.4b	*C. quercinus*			P4,P2		DG**CC**PNPS**C**SVNNPDI**C**[a]	α4/7	DQ311069
Qc1.5	*C. quercinus*			P4,P2		G**CC**SNPA**C**SVNHPEL**C**[a]	α4/7	DQ311070
Qc1.6	*C. quercinus*			P4,P2		G**CC**SNPT**C**AGNNGNI**C**[a]	α4/7	DQ311071
Pu1.1	*C. pulicarius*			P1,P2	1167	QN**CC**NVPG**C**WAKYKHL**C**[a]	α4/7	DQ309775
				P4,P2				DQ359142
Pu1.2	*C. pulicarius*			P4,P2		GG**CC**SYPP**C**IANNPL**C**	α4/6	DQ311078
Pu1.3	*C. pulicarius*			P4,P2		LS**CC**ADPA**C**KHTPG**C**	α4/5	DQ311079

[a]C-terminal amidation (Buczek et al., 2005).

peptide of 21 residues, a pro-region of about 27 residues and a hypervariable mature peptide. This precursor structure is also shared by the conotoxins of other superfamilies, such as M-, T-, O-superfamily (Olivera, 1997; Terlau and Olivera, 2004). This implies that all these conotoxins experience the similar processing procedure after being translated, the cleavage of the signal peptide followed by the cleavage of the pro-peptide. Another processing step is the post-translational modification which is quite common in conotoxins (Craig et al., 1999; Buczek et al., 2005), but it is not clear yet this step is before or after the cleavage of pro-peptide. An exception is the precursor of κ-BtX which contains no pro-peptide but a 13 amino acid post-peptide instead (Fan et al., 2003). Recently, this kind of

post-peptide has been found to be the recognition site for the gamma-carboxylation of glutamyl residue (Brown et al., 2005).

In order to get cDNAs of α conotoxins with full length 3′UTR, forward primer P1 was paired with an abridged universal amplification primer P3 for PCR amplification. Five α conotoxin cDNAs were identified by this way (Table 1). Among them, Qc1.1a, Qc1.1b and Lp1.1 have already been characterized with primers P1 and P2. Four of them (Qc1.1a, Qc1.1b, Lp1.7, Lp1.8) possess a relatively long 3′ UTR of 658–820 bp; however, the 3′UTR of Lp1.1 is only 225 bp. Sequence comparison (Fig. 2) suggests that the significant difference of 3′UTR length resulted from a large 3′ end deletion in Lp1.1 cDNA, although the major part of its 3′UTR is quite well aligned with the 3′UTR sequences of other α conotoxins. No hint could be found yet to explain or predict this large 3′ end deletion of Lp1.1. The only striking point of Lp1.1 is that the third residue in the first Cys loop of mature toxin is not the conserved Pro, but an Ala. It is not clear whether this is related with its notably short 3′UTR.

The sequence comparison also clearly indicates that the 3′ UTR of Lp1.7 and Lp1.8 are very different from the sequences of other α conotoxins (Fig. 2). This might be the reason why the cDNAs of these two toxins were not obtained when conserved α conotoxin 3′ UTR specific primer P2 was used for cloning. Their toxin sequences with a rather long N-terminal part prior to the two adjacent Cys residues are also quite different from those of other α4/7 conotoxins, and their pro-region is obviously shorter. The second half sequences of their precursors, encoded by exon II (see below, Table 2), are very different from the sequences of other α conotoxins, too. The sequence of Lp1.7 is completely the same as that of Lt1c, a library clone from *C. litteratus* (Genbank no. DQ345366), a conus species looking very much like and probably close related with *C. leopardus*. These conotoxins might define a new subfamily within the α conotoxins.

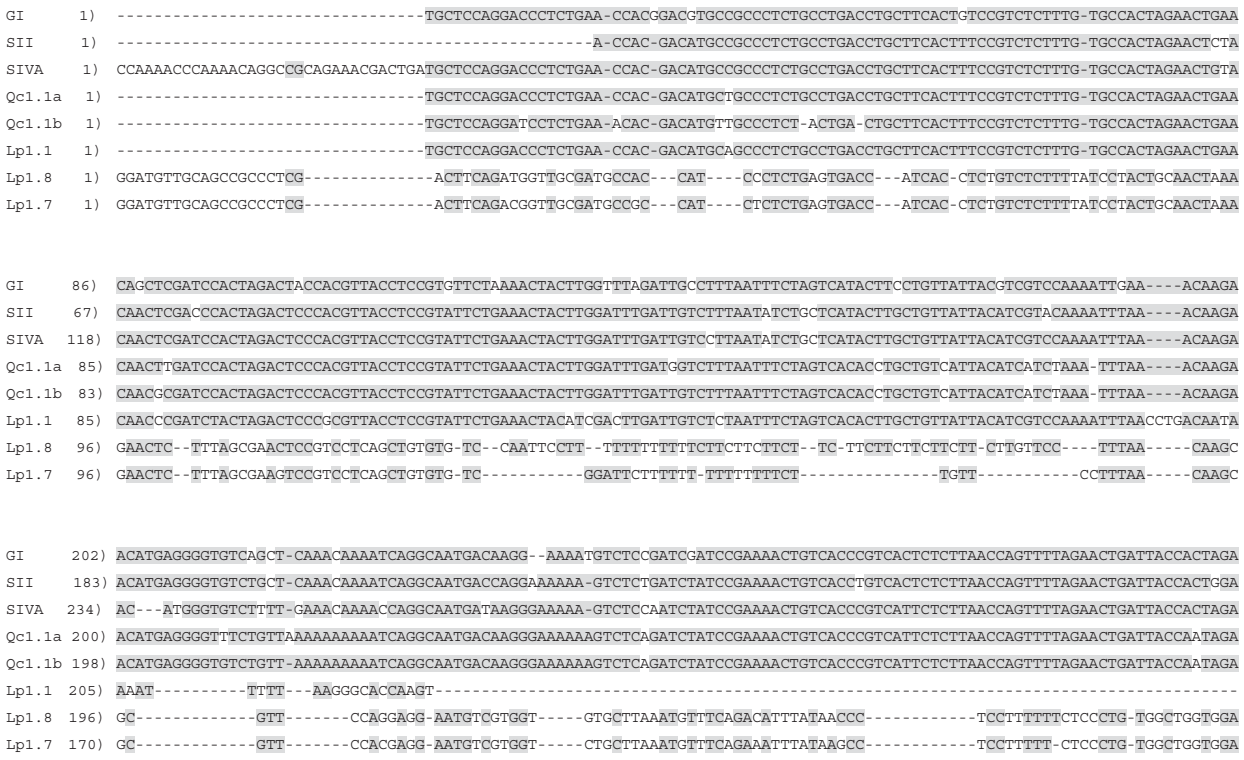

Fig. 2. Sequence comparison of 3′ UTR of α conotoxins. The conserved nucleotides are shadowed and gaps (–) have been inserted to maximize similarity. 3′ UTR sequences of Qc1.1a, Qc1.1b, Lp1.1, Lp1.7 and Lp1.8 are from this work, and sequences of GI, SII and SIVA are from references (Wang et al., 2003; Santos et al., 2004).

```
GI     319) GCTTTTGTACCACATCAAATCAGGTCTATGTGTGATGTTTC-TTTTGCAAA-----ATTTAATTTTTGAGAAAAAAA-GCTCAAAATGTGGGAAGTGCTTTTGATTTTCTGACAACTTGT
SII    301) GCTTTTGTACCACATC-AACCAGGTCTATGTGTAATGTTTCGTTTTGCAAACTTTGCTGTAATTCTTGACAAAAAA--GCTCAAAATGTGGGAAGCGCTTTTGATTGTCTGACAACTTGT
SIVA   349) GCTTTTGTACCTCATCAAATCCGGTCTATGTGTAATGTCT--CTTTGCAAACTTCGCTGAAATTCTTGAGAAAAAAA-GCTCAAAATGTGGGAAGTGCTTTTGGTTATCTGACAACTCGT
Qc1.1a 320) GCTTTTGTACCACATAAAATCAGGTCTTTGTGTAATGTTTC-TTTCACAAACGTCGCTGAAATTCTTGGGAAAAAAT-GCTCAAAATGTGGGAAGCACTTTTGGTTGTCTGACAACTTGT
Qc1.1b 317) GCTTTTGTACCACATAAAATCAGGTCTTCGTGTAATGTTTC-TTTTACAAACGTCGCTGAAATTCTTGGGAAAAAAT-GCTCAAAATGTGGGAAGCACTTTTGGTTGTCTGACAACTTGT
Lp1.1  226) ----------------------------------------------------------------------------------------------------------------------
Lp1.8  277) -TTGTGATGTCGCTTA---------ATTCGTCTCAACATTG-----TTAAATTCTGCTTCGATTTGAGATTCCAATTATAGCAATACTTGGATATGACATATCATCCTCGGAAAAGGCCA
Lp1.7  250) -TTGTAATGTCGCTTA---------ATTCGTCTCAACATTG-----TTAAATTCTGCTTCGATTTGAGTTTCGAATTATAGCAATACTTGGAAATGACATATCATCCTCGGAAAAGGCCA

GI     432) GATCATGTCCGTTTTCAGTGAGTCTAATTGCAACCTCTGTGTGATTTTCTTCACCTGTTAAGCAACGCAAAGAGGTTGTCCATAACCAGGAAAGCAACAGACAAAGAAATGCTTGAGAAT
SII    418) GATCATGTCTCTTTTCAGTGAGTCTAAGTGCAACCTCTGTGTGTTTTCCTTCACCTGTTAAGCAACGAAAAGAGGTTGTCCATAATCAGAAAAGCAAAAGACAAAGAACTGCTTGAGAAT
SIVA   466) CATCATGTTCTTTTTCAGTGAGTCTAACTGCAACTTTGTGTCTTTTTCTTCACTTGGTAAGCAACACAAA---GTTGTCCATAATCAGGAAAGCAAAAGATAAAGAACTGTTTGAGA-T
Qc1.1a 438) GATCATGTCTCTTTTCAGTGAGTCTAACTGCAACCTCTGTGTGTTTTCCTTCACCTGTTAAACAATGCAAAGTGGTTTTCCATAATCAGAAAAGCCAAAGACAAAGAACCGCTCGATAAT
Qc1.1b 435) GATCATGTCCCTTTTCAGCGAGTCTAACTGCAACCTCTGTGTGTTTTCCTTCACCTGTTAAACAATGCAAAGTGGTTTTCCATAATCAGAAAAGCCAAAGACAAAGAACCGCTCGATAAT
Lp1.1  226) ----------------------------------------------------------------------------------------------------------------------
Lp1.8  382) ATTGATGTCAATTTT----------AGCGGGAAAGAGAGAGATAGAAAAAAAAAAAAG-GAAGTTACGGAACCTTGAAAAACAGACAAAGACAA--AACAGGAAAACAA-------ACTTT
Lp1.7  355) ATTGATGTCCATTTT----------AGCGGGAAAGACAGAG----ATAGAAAAAAAAG-GAAGTTCCGGAACCTTGGAAAACAGACAAAGACAA--AACAGGAAAACAA-------ACTGT

GI     552) TTCAGGTTATAGATAAGGTAAGGAA---AAAAAGGAGAGCTATGGGAAATGATGAAAACAACAGATAAAATAAATTGAACAGTACCTACTTGTTTCATGGTTGATTT-TTTTTT-CTCTG
SII    538) TTCAGGTTATATATAGGGTAAGCGATAAAAAAAGCACAGCTGTGAGAAATGATGAAAACAACAGATAAAAGAAATTGAACAATACCTACTTGTTTCATGGTTGATGA-TTCTTT-TTT--
SIVA   582) TTCAGGCTATAGATAGGGTAAGG-----AAAAACAGAGCTATGGGAAATGGTGAAAACAACAGATAAAAGAAATTTAACAATAACTACTTGTTTCATGGTTGATGA-TTTTTTGTTCTG
Qc1.1a 558) TTCAGGAAATAGATAGGGTAAGGAA----AAAAAGAGACCTATGGGAAATGATAAGAACAACAGATAAAAGAAATTGAACATTACCTACTTGTTTCATGGTTGATGACTTTTTTTTTCTG
Qc1.1b 555) TTCAGGAAATAGATAGGGTAAGGGA----AAAAAGAGACCTATGGGAAATGATAAGAACAACAGATAAAAGAAATTGAACATTACCTACTTGTTTCATGGTTGATGACTTTTTT-TTTT-
Lp1.1  226) ----------------------------------------------------------------------------------------------------------------------
Lp1.8  482) TCTCGGTCGTAGTTT---GTTTGCTT---------------TCTTGTACCCCTTAAAAA-------GAAAAGAAA--GAAAATTATCAGTCTGGATCA--GCT-ATCT-TAATCTGTTCT-
Lp1.7  451) TCTTGGTAGTAGTTT---TTTTGCTT---------------TCTTATTCCCGTTAAAAA---AAAAAAAAAGAAA--GAAAATTAACAGTCTGGATTA--GCT-ATCT-TAGTCTGTT---

GI     667) AATAATCTCTGTGGACACTAATGGCAGTCTCTCCTCACCCC-ACGCCATTAG------TAAGCTTATTTTTTCTTTCTTT----ATCCAAGATTTGCTGAACATATTTAGCCTAGATATA
SII    654) AATAATCTTTGTAGACAGTAATGGCAGTCTCTC-TCCCCCC-ATGCCATT---------------TTTTTTCTTTCTTT----ATTCAAGACTTGCTGAATA---TTA----------
SIVA   695) AATAATGTTTGTGGATATGAATGGCAGTGTCTCTTCCCCCCAATGCCATTAGTTTTTTTTTTTTTTTTTTTGTCTTTCTTTCTTTATCCAAGATTTGCTGAATA---TTA----------
Qc1.1a 674) AATAATCTCTGTGGACATTAATTGCAGTCTCTCCTCATCCC-ATGCCATTAG------------TTTTTTCTTTCTTT----ATCCAAGATTTGCTGAATATTTTTAGCCTAGATATA
Qc1.1b 669) AATAATCTCTGTGGACATTAATTGCAGTCTCTCCTCATCCC-ATGCCATTAG------------TTATTTTTCTTTCTTT----ATCCAAGATTTGCTGAATATTTTTAGCCTAGATATA
Lp1.1  226) ----------------------------------------------------------------------------------------------------------------------
Lp1.8  570) CATCCACTTTAT--ATAGGGAT---AATATTTC------------TATTTG-----------ATTGCATTTATTCATT-TTAATCTGACCATTGCTGAAGATTCTTTGCC------TA
Lp1.7  541) -ATCTACTTTAT--ATAGGGAT---ATTATTTC------------TATTTG-----------ATTGCATTTATTCATT-TTAATCTGACCATTGCTGAAAATTCTTTGCC------TA

GI     776)  GACATTGCTACATATATAATCTGACAATAAAC-----TTTCATGGGCACCAATT-poly(A)
SII    738)  ----------------TGAACTGAC-ATAAAC-----TTTCATGGGCACCAATTCpoly(A)
SIVA   801)  ----------------TGAACTGACAATAAAC-----TGTCATAGGCACC-----poly(A)
Qc1.1a 776)  GACATTGCTACATGTATAACCTGACAATAAAC-----TTTCATGGGCACC-----poly(A)
Qc1.1b 772)  GACATTGCTACATGTATAACCTGACAATAAAC-----TTTCATGGGCACCAATT-poly(A)
Lp1.1  226)  ----------------------------------------------------poly(A)
Lp1.8  653)  TAC------ACA----TGACCTGACAATAAACATTTTTTGGAGGTGCA-------poly(A)
Lp1.7  623)  TAC------ACA----TGACCTGACAATAAACATTTTTTGGATGTG---------poly(A)
```

Fig. 2. (*Continued*)

3.2. Gene cloning of α conotoxins

From the genomic DNAs of five species of cone snails, partial gene sequences of six α conotoxins (Mr1.3, Pu1.1, Lp1.1, Lp1.4, Lp1.6a and Ac1.1b) plus two pseudogenes (Qc αL-1 and Qc αL-2) were obtained (Fig. 1, Table 1). In these pseudogenes the first cysteine residue in mature peptides are mutated into phenylalanine due to a point mutation (TGC to TTC). Further PCR experiment with a different primer set has proven this mutation is not an error of PCR.

As the cDNAs of Lp1.1, Lp1.4, and Ac1.1b have already been characterized, the comparison of the cDNA sequences and the corresponding genomic sequences allowed us to assign unambiguously the

Table 2
Comparison of the prepropeptide sequences of α conotoxin

Conotoxin	Identity in exon I (%)	Prepropeptide sequence	Identity in exon II (%)
A. conotoxins cloned from cDNA			
Mr1.1	100	MGMRMMFTVFLLVVLATTVVSFTSDRASDGRKAAAKDKASDLVALTVKGCCSHPACSVNNPDICG	100
Mr1.2	100	MGMRMMFTVFLLVVLATTVVSFTSDRGSDGRNAAAKDKASDLVALTVKGCCSNPPCYANNQAYCNGRR	78
Qc1.1a	88	MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAANDKASNLMALRDECCPDPPCKASNPDLCDWRS	56
Qc1.1b	88	MGMRMMFTMFLLVVLAITVVSFTSDHASDGRNTAANDKASKLMALRNECCDNPPCKSSNPDLCDWRS	56
Qc1.2	88	MGMRMMFTVFLLVALATTVASFTLDRASNGRNAAADDKPSDWIALAIKQCCANPPCKHVNCR	58
Ac1.1a	81	MGMRMMFTLFLLVVLTTTVVSYPSDSASDGRDDEAKDERSDMYELKRNGRCCHPACGKHFNCGR	55
Lp1.7	88	MGMRMMFTMFLLVVLTTTVVSFNSDRESNHENRRTSNQITRGMWDECCDDPPCRQNNMEHCPAS	21
Lp1.8	88	MGMRMMFTMFLLVVLTTTVVSFNSDRESNHENRRTSNQITRGVWDECCKDPQCRQNHMQHCPAR	19
B. conotoxins cloned from genomic DNA			
Mr1.3	92	MGMRMMFTMCLLVVLATTVISFTSDRASNGRNAAAKDKASDLNALNVRGCCSHPACRVHYPHVCYGRR	72
Pu1.1	100	MGMRMMFTVFLLVVLATTVVSFTSDRTSDGRNAAFNAFDLIALTARQNCCNVPGCWAKYKHLCGRKR	46
Lp1.6a	85	MGMRMMFIIFLFVVLATTVVSFTSGRASDGRNAPANNKVSDLIRQFCCGHYDCDFIPNVCG	44
Qc αL-1	81	MGMRMMFTMFLLVVLATTVVSINLDHAFDGRNAAANNKATDLMARTVRRFCSDPPCRISNPESCGWEP	56
Qc αL-2	85	MGMRMMFTMFLLVVLATTVVSFNLDHAFDGRNAAANNKATDLMARTVRGFCSDPSCRFGNPELCDWRR	58
C. conotoxins cloned from both cDNA and genomic DNA			
Ac1.1b	85	MGMRMMFTVFLLVVLTTTVVSYPSDSASDGRDDEAKDERSDMYESKRNGRCCHPACGKHFSCGR	51
Lp1.1	77	MGMRMMFIMFMLVVLATTVDTFTSDHALDAMNAAASNKASRLIALAVRGCCARAACAGIHQELCGGRR	48
Lp1.4	88	MGMRMMFIMFMLVVLATTVVSFTSDRALDAMNAAASKKASRLIALAVRGCCSHPACSGNHQELCDGRR	61

The signal sequence is underlined and the toxin region is shadowed. The intron inserting residues are boxed.

exon/intron boundary and structural organization of these genes. The cDNA and gene sequences of Lp1.1 from *C. leopardus* are shown as an example in Fig. 3. It appears that all these three α conotoxins share the same gene organization, namely, two exons interrupted by a rather long intron of more than 700 bp. Exon I encodes the signal peptide and the first several amino acids of the propeptide, while exon II encodes the other part of propeptide, the whole mature peptide and 3′ UTR. Additionally all these introns are inserted at the same position, namely in the codon of the 6th residue of the pro-region, which is Ala in most cases (Fig. 3, Table 2). Sequence analysis of the other five partial gene sequences (Mr1.3, Pu1.1, Lp1.6a, Qc αL-1 and Qc αL-2) strongly suggest that they may possess the same gene structure, which is also very similar with that of conventional short-chain and long-chain scorpion toxins although the introns of α conotoxin genes are significantly longer. This gene structure, namely the interruption of an intron, is probably required for the gene recombination and evolution of conotoxins.

Although these introns vary from 791 bp (Lp1.1) to 1441 bp (Qc αL-2) in length, they have some common features. They are all flanked by a gt/ag donor–acceptor pair, which is consistent with all the introns in eukaryotic genes. However, the AT content in the introns of α conotoxins is around 60%, significantly lower than that of scorpion toxins (Xu et al., 2004; Wang et al., 2005). And the sequences near the splice donor/acceptor sites in all these introns are conserved except the 5′ end sequence in the intron of Lp1.1 (Fig. 4). The most remarkable feature of these introns is their highly conserved 3′ end sequence, the last 14 nucleotides of introns being completely identical (-ggttctgggtccag-) and preceded by a stretch of GT dinucleotide repeats of different length. The high conservation of the 5′ and 3′ end of these introns may suggest a specific, but unknown yet, intron splicing mechanism during transcription.

Another notable feature of these introns is the existence of two kinds of repetitive sequences. The first one is stretches of dinucleotide or trinucleotide repeats "GT", "GA", "CA", "CAT" in the middle parts of introns (Fig. 4). The number of this oligonucleotide repeats ranges from 13 to over 40. And there is unexceptionally a tract of GT repeats preceding the last 14 conserved nucleotides in 3′ end of all the introns (Fig. 4). As known, in contrast to prokaryotic genomes, eukaryotic genomes possess a large amount of different kinds of simple tandem repeats which is regarded as "micro-satellite DNA" (Debrauwere et al., 1997) and may carry out some regulatory functions. For example, a stretch of "GT" repeat in the 5′ sequence of intron 2 of human Na^+/Ca^{2+} exchanger 1 gene (*NCX1*) is an intronic splicing enhancer involved in regulation of *NCX1* expression (Gabellini, 2001). And the $(TG)_{12}$ repeat in the promoter region of venom group I phospholipase A_2 genes from cobra *Naja sputatrix* was found to act as a silencer in gene regulation (Jeyaseelan et al., 2000). For conotoxins, further study is needed to elucidate the actual functions of these repeats.

Another kind of repeat was found only in the intron of α conotoxin Pu1.1. In contrast to the above short sequence repeat of DNA, the unit of DNA repeat in Pu1.1 is about 140–150 bp in length. There are five such repeats, and the level of conservation in these repeats is decreasing from 5′ to 3′ end (Fig. 5). This makes the intron of Pu1.1 not well aligned with other intron sequences except the 5′ end (Fig. 4). Such DNA repeats of comparatively long sequence have also been characterized in the intron of snake venom group IA phospholipase A_2 (Fujimi et al., 2002). But the function of this long repeat is elusive.

The comparison of all the eight intron sequences of α conotoxins (Fig. 4) showed that intron is more conserved within species, for example, between Lp1.1, Lp1.4 and Lp1.6a, or between QcαL-1 and QcαL-2. But between species, large gap could be found. A gap of about 500 bp could be found after the 5′ end 70 bp in the introns of toxins from *C. leopardus*, and for intron of Mr1.3, a gap of about 600 bp is found at 3′ end, just before the conserved GT repeats and identical 3′ end 14 bp. These gaps result in the intron length difference. Apart from these, the introns of two pseudogenes, QcαL-1 and QcαL-2, are quite similar with the intron of Ac1.1b, showing no special feature for pseudogene.

3.3. Exon II cloning of α conotoxins

The extraordinarily high conservation in 3′end of introns in α conotoxin genes enables us to clone novel α conotoxins genes from genomic DNA directly. A forward primer P4 which matches the conserved 3′ end sequence of introns, together with reverse primer P2, was used for PCR amplification. By this way, exon II of 16 α conotoxin and a pseudogene Qc αL-1 were identified from genomic

a

```
    ATG GGC ATG CGG ATG ATG TTC ATC ATG TTT ATG TTG GTT GTC TTG GCA ACC ACT GTC
1     M   G   M   R   M   M   F   I   M   F   M   L   V   V   L   A   T   T   V

    GTT  ACC TTC ACT TCA GAT CGT GCA CTT GAT GCC ATG AAT GCT GCA GCC AGC AAC AAA
20    V   T   F   T   S   D   R   A   L   D   A   M   N   A   A   A   S   N   K

    GCG TCT  CGC CTG ATC GCC CTG GCC GTC AGG GGA TGC TGT GCC CGT GCT GCC TGT GCC
39    A   S   R   L   I   A   L   A   V   R   G   C   C   A   R   A   A   C   A

    GGG ATT CAT CAA GAA CTT TGT GGA GGA GGA CGC TGA  tgctccagga ccctctgaac
58    G   I   H   Q   E   L   C   G   G   G   R  End

  cacgacatgc agccctctgc ctgacctgct tcactttccg tctctttgtg ccactagaac tgaacaaccc
  atctactaga ctcccgcgtt acctccgtat tctgaaacta catcgacttg attgtctcta atttctagta
  cacttgctgt tattacatcg tccaaaattt aacctgacaa taaaattttt aagggcacca agt poly(A)
```

b

```
    ATG GGC ATG CGG ATG ATG TTC ATC ATG TTT ATG TTG GTT GTC TTG GCA ACC ACT GTC
1     M   G   M   R   M   M   F   I   M   F   M   L   V   V   L   A   T   T   V

    GAT ACC TTC ACT TCA GAT CAT G/gtcagtctgc atttgttaag tcaaaatccg aattctcatg
20    D   T   F   T   S   D   H
  ttttgtaacc tggttagctt tgagacgtag tgtactatca ccatcatcat catcatcatc atcatcatca
  tcatcatcat catcatcaac atcatcatca ttaccactag tattataagt aattatagtt gaagtaatag
  gagtagtagt ggtagtagta cttatatatt gcttgtcttc agttaaagac caggttcaga gtatttcaca
  atcaacgcta gcatttcaac ataaggctgc ctgcttgggt agagcagcca aagggtcgct tactggcgct
  tcattaagtt gttttcagtg tcgttcagtt actattaaac attctgttct tttattaaca tttttagatt
  ttttttcacg tagtgcgcgt actcatcatc atcatcatca tcatcatcat catcatcatc atcatcatta
  ttactaatag tagtagaagt agttgtagtt gtagtaatag aaatagtggt ggtagcagta cttatgtttc
  acaaggtgcc catttagatt cttgtcatcc agtgcgcgta ctatccatac caacgaccaa actagccttc
  tcttctttac gttgatgatg gggtttgaca tctatctaag ctgttttgtc tgtctgtctg tctgtctgtc
  tatctttcca tctgtctgtt cgtgagagag agagagaggg agagtacatg tgggtgtacg cgcgcgcgtg
  tgtgtgtgtgtgtgtgtgtg tgtgtgtgtg tgtgtgtggt tctgggtccag /CA CTT GAT GCC ATG AAT
                                                          A   L   D   A   M   N
    GCT GCA GCC AGC AAC AAA GCG TCT CGC CTG ATC GCC CTG GCC GTC AGG GGA TGC TGT
33    A   A   A   S   N   K   A   S   R   L   I   A   L   A   V   R   G   C   C

    GCC CGT GCT GCC TGT GCC GGG ATT CAT CAA GAA CTT TGT GGA GGA CGA CGC TGA
52    A   R   A   A   C   A   G   I   H   Q   E   L   C   G   G   R   R  End

  tgctccagga ccctctgaac cacgac
```

Fig. 3. cDNA (a) and partial gene (b) sequences of Lp1.1 with the predicted translation product. The signal sequences are shaded, and the toxin region is underlined. Nucleotides in the untranslated region are in small letters. The dinucleotide repeats in intron are indicated by shadow. The poly(A) addition signal (aataaa) and the intron boundary gt/ag are also indicated. There are three single nucleotide polymorphism between the cDNA and gene sequences, but out of the mature peptide encoding region.

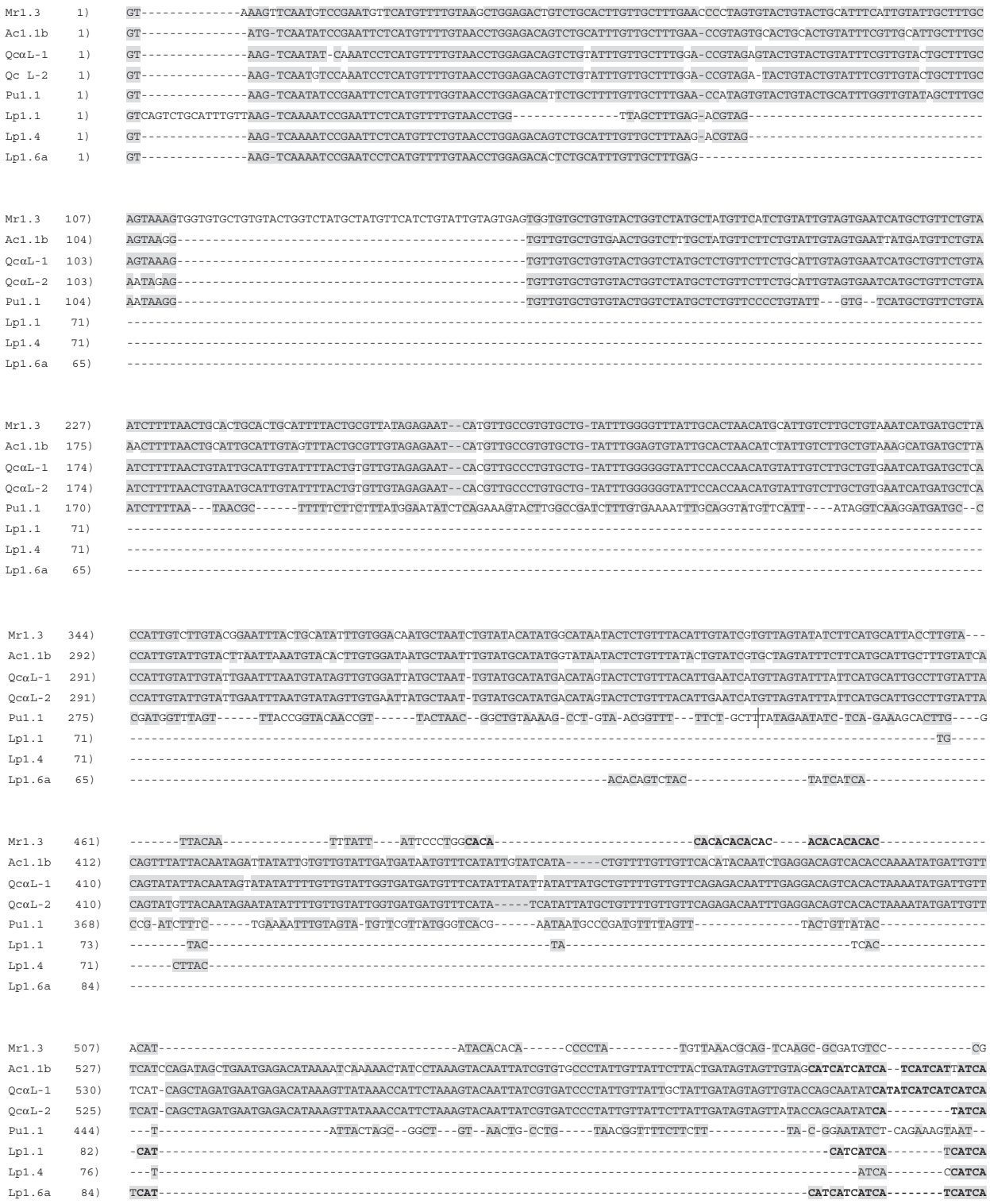

Fig. 4. Sequence alignment of eight introns of α conotoxins. The conserved nucleotides are shadowed and gaps (–) have been inserted to maximize similarity. The completely identical 3′ end sequences are underlined. The dinucleotide or trinucleotide repeats are in bold.

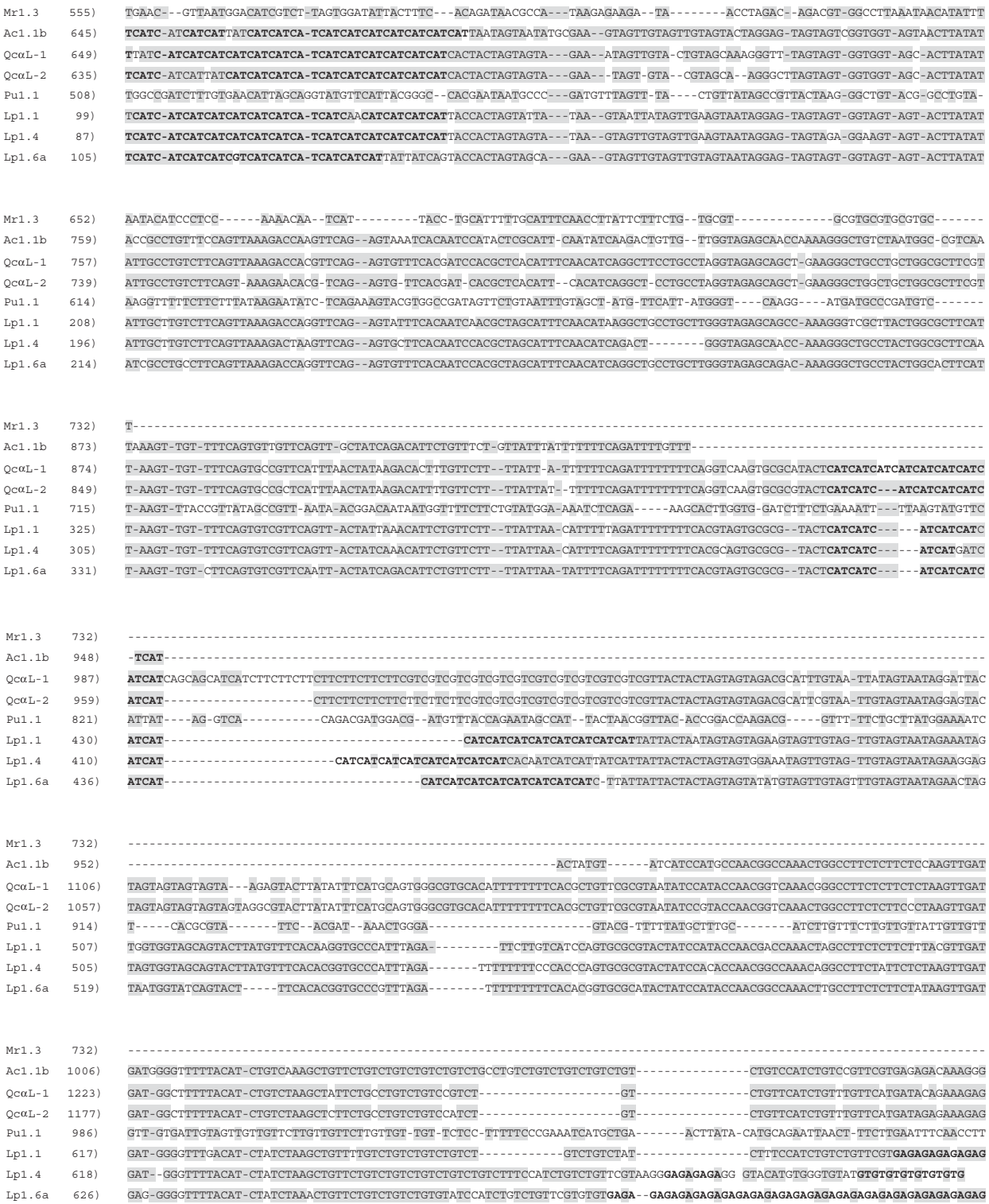

Fig. 4. (*Continued*)

Fig. 4. (*Continued*)

Fig. 5. Sequence alignment of five DNA repeats found in the intron of Pu1.1. The numbers indicate the location of the repeats in the intron. The conserved nucleotides are shadowed and gaps (–) have been inserted to maximize similarity.

DNAs of *C. marmoreus, C. quercinus*, *C. leopardus, C. pulicarius* and *C. achatinus* (Fig. 1, Table 1). Some of them, such as Mr1.1, Lp1.6a, Pu 1.1, Ac1.1b and Qc αL-1, have already been cloned from genomic DNA or cDNA. Among these 16 α conotoxins, there are one α3/5, one α4/4, three α4/5, one α4/6 and 10 α4/7 α conotoxins. This strategy has been used to clone α conotoxin GIC and PIA (McIntosh et al., 2002; Dowell et al., 2003). Since it is impossible for cone snails to express all the conotoxins in arsenals at the same time, such cloning from genomic DNA is a better alternative to acquiring more conotoxins than cDNA cloning.

Unexpectedly, with primers P4 and P2 we also got three novel κA conotoxins genes with a totally different Cys pattern (-CC-C-C-C-C-) from the piscivorous cone snail *C. achatinus* (Table 1). They apparently share the same conserved intron 3′end and 3′UTR sequence of α conotoxins, which strongly supports the point that κA conotoxins and α conotoxin belong to the same A superfamily (Santos et al., 2004).

3.4. A hypothesis of α conotoxin gene evolution

Low mobility and extreme variety of prey (five different phyla) force cone snails to produce hypervariable toxins bearing the ability to target different subtypes of the same membrane receptor or ion channel with high affinity. Numerous cDNA cloning studies, including this work, have shown that each individual conotoxin has its own gene and that these genes encode the similar prepropeptide precursor. This indicates that whole gene duplication in the same gene family, followed by hypermutation in mature peptides, is the major way at the gene level to get more variable toxins (Kordis and Gubensek, 2000; Sollod et al., 2005). The DNA sequences we got from five conus species of all three feeding types (fish-hunting: *C. achatinus*;

mollusk-hunting: *C. marmoreus*; worming-hunting: *C. quercinus*, *C. leopardus, C. pulicarius*) give some hints about how this hypermutation is achieved.

Sequence analysis clearly showed that the variation in exon II is significantly higher than that in exon I (Table 2). This may be related with the gene structure of α conotoxins, which was firstly clarified in this study. As suggested previously (Olivera et al., 1997), interruption by the long intron makes it possible that the two exons undergo different control in DNA replication, exon II being "encouraged" to mutate by looser surveillance. However, the N-terminal part of exon II encoded region, i.e. the middle part of the pro-peptide, is relatively conserved, which might be the result of the natural selection pressure caused by the requirement of the essential function of pro-peptide. The pro-peptide of an α conotoxin has been found to play a role in the PDI (protein disulfide isomerase)-catalyzed oxidative folding of conotoxin precursors (Buczek et al., 2004).

In exon II, the point mutation is probably the basic mechanism to have different toxins, such as Ac1.1a vs. Ac1.1b, Qc1.4a vs. Qc1.4b. Then the accumulation of point mutations can give more varied conotoxins with the same length and framework, like Qc1.5 and Qc1.6. In these cases, it is hard to tell which toxin appears earlier and which one follows after mutation. But for different spacing α conotoxins, it is proposed that α4/7 subfamily is the ancestor since this group distribute in all the conus species; other subfamilies evolved from α4/7 toxins (Santos et al., 2004). The gene sequences indicate that one or more codons are deleted to have conotoxins with different loop spacing, for instance, deletion of the Asp codon from the second loop of α4/7 Qc1.4a, together with some point mutations, could give an α4/6 conotoxin Pu1.2. Similar is the case from α4/5 Pu1.3 to α4/4 Qc1.2. The mechanism of the codon deletion still remains to be clarified.

The sequences of two pseudogenes Qc αL-1 and Qc αL-2 also suggest another possible evolution pathway to have α conotoxins of different Cys framework (Fig. 6). With the first Cys (TGC) mutated into Phe (TTC), the pseudogenes with odd Cys numbers must be the intermediate product of evolution process. Most likely, the next step is the mutation of the first codon (TCC) in the first loop into another Cys (TGC). Thus, the framework CCX_4CX_7C would be shifted into $FCCX_3CX_7C$, then deletion of two codons in the second loop would give an α3/5 conotoxin. The unusually high mutation rate of conotoxin genes (Kordis and Gubensek, 2000; Sollod et al., 2005), as well as the various gene block deletion or substitution (Santos et al., 2004), makes these steps highly possible.

An alternative α4/7 to α3/5 evolution pathway (CCX→CCC→XCC) has been proposed recently, the supporting evidence of which is the nonregular six-Cys-containing α conotoxin SII (Santos et al., 2004). But each step of this pathway needs to be accompanied with the addition or the deletion of the C-terminal fragment including the 6th Cys, which is not easier than the hypothesis supported by these pseudogenes (CCX→FCX→XCC). More evidence is certainly needed to distinguish which one is more possible.

It would be intriguing to clarify whether the corresponding three-Cys-containing conotoxins of these pseudogenes are produced or not and how these toxins, if produced, fold and act. The in vivo maturation process is believed to deliver the conotoxins into an oxidative environment, where the thiol group of Cys residue is unlikely free. The unpaired Cys would possibly be linked with a small oxidizing/reducing molecule, such as glutathione.

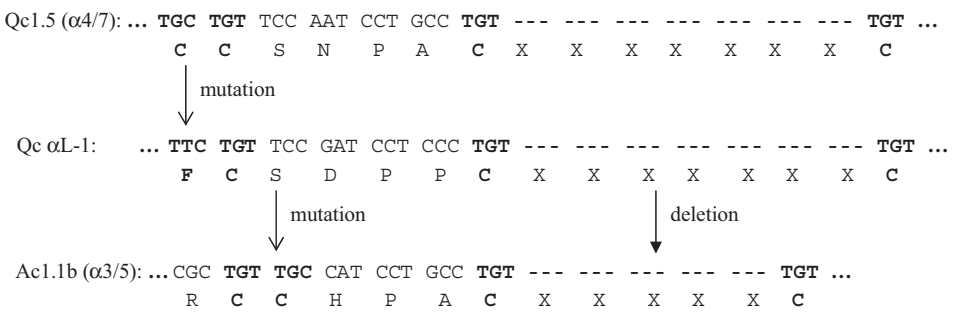

Fig. 6. Hypothesis of the gene evolution from α4/7 to α3/5 toxin. The mutation of the first Cys codon might be followed by the mutation of the first codon in the first loop into a new Cys, by which the CCX_4CX_7C framework could be shifted into $XCCX_3CX_7C$ framework. Then deletion of two codons in the second would give a α3/5 framework.

Another possibility which cannot be ruled out is that the unpaired Cys of one toxin might form a disulfide bond with the counterpart Cys of another toxin to get stabilized.

In summary, based on the conserved signal peptide and 3′UTR sequence, 16 novel α contoxins were identified from cDNA and genomic DNA. The gene structure and sequence further enable us to get other 12 novel α conotoxin sequences by using exon II cloning. These sequences will certainly enrich our knowledge about the α conotoxin diversity as well as the evolution process.

Acknowledgments

This work was supported by the National Basic Research Program of China (2004CB719904). Chunguang Wang is supported by the Program for young excellent talents in Tongji University (2006KJ063) and Dawn Program of Shanghai Education Commission (06SG26).

References

Azam, L., Dowell, C., Watkins, M., Stitzel, J.A., Olivera, B.M., McIntosh, J.M., 2005. Alpha-conotoxin BuIA, a novel peptide from *Conus bullatus*, distinguishes among neuronal nicotinic acetylcholine receptors. J. Biol. Chem. 280, 80–87.

Brown, M.A., Begley, G.S., Czerwiec, E., Stenberg, L.M., Jacobs, M., Kalume, D.E., Roepstorff, P., Stenflo, J., Furie, B.C., Furie, B., 2005. Precursors of novel Gla-containing conotoxins contain a carboxy-terminal recognition site that directs gamma-carboxylation. Biochemistry 44, 9150–9159.

Buczek, O., Olivera, B.M., Bulaj, G., 2004. Propeptide does not act as an intramolecular chaperone but facilitates protein disulfide isomerase-assisted folding of a conotoxin precursor. Biochemistry 43, 1093–1101.

Buczek, O., Bulaj, G., Olivera, B.M., 2005. Conotoxins and the posttranslational modification of secreted gene products. Cell Mol. Life Sci. 62, 3067–3079.

Cartier, G.E., Yoshikami, D., Gray, W.R., Luo, S., Olivera, B.M., McIntosh, J.M., 1996. A new alpha-conotoxin which targets alpha3beta2 nicotinic acetylcholine receptors. J. Biol. Chem. 271, 7522–7528.

Craig, A.G., Bandyopadhyay, P., Olivera, B.M., 1999. Post-translationally modified neuropeptides from Conus venoms. Eur. J. Biochem. 264, 271–275.

Debrauwere, H., Gendrel, C.G., Lechat, S., Dutreix, M., 1997. Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. Biochimie 79, 577–586.

Dowell, C., Olivera, B.M., Garrett, J.E., Staheli, S.T., Watkins, M., Kuryatov, A., Yoshikami, D., Lindstrom, J.M., McIntosh, J.M., 2003. Alpha-conotoxin PIA is selective for alpha6 subunit-containing nicotinic acetylcholine receptors. J. Neurosci. 23, 8445–8452.

Fan, C.X., Chen, X.K., Zhang, C., Wang, L.X., Duan, K.L., He, L.L., Cao, Y., Liu, S.Y., Zhong, M.N., Ulens, C., Tytgat, J., Chen, J.S., Chi, C.W., Zhou, Z., 2003. A novel conotoxin from Conus betulinus, kappa-BtX, unique in cysteine pattern and in function as a specific BK channel modulator. J. Biol. Chem. 278, 12624–12633.

Favreau, P., Krimm, I., Le Gall, F., Bobenrieth, M.J., Lamthanh, H., Bouet, F., Servent, D., Molgo, J., Menez, A., Letourneux, Y., Lancelin, J.M., 1999. Biochemical characterization and nuclear magnetic resonance structure of novel alpha-conotoxins isolated from the venom of *Conus consors*. Biochemistry 38, 6317–6326.

Fujimi, T.J., Tsuchiya, T., Tamiya, T., 2002. A comparative analysis of invaded sequences from group IA phospholipase A(2) genes provides evidence about the divergence period of genes groups and snake families. Toxicon 40, 873–884.

Gabellini, N., 2001. A polymorphic GT repeat from the human cardiac $Na+Ca_2+$ exchanger intron 2 activates splicing. Eur. J. Biochem. 268, 1076–1083.

Gray, W.R., Luque, A., Olivera, B.M., Barrett, J., Cruz, L.J., 1981. Peptide toxins from *Conus geographus* venom. J. Biol. Chem. 256, 4734–4740.

Han, Y.H., Wang, Q., Jiang, H., Miao, X.W., Chen, J.S., Chi, C.W., 2005. Sequence diversity of T-superfamily conotoxins from *Conus marmoreus*. Toxicon 45, 481–487.

Jeyaseelan, K., Armugam, A., Donghui, M., Tan, N.H., 2000. Structure and phylogeny of the venom group I phospholipase A(2) gene. Mol. Biol. Evol. 17, 1010–1021.

Kordis, D., Gubensek, F., 2000. Adaptive evolution of animal toxin multigene families. Gene 261, 43–52.

Loughnan, M., Bond, T., Atkins, A., Cuevas, J., Adams, D.J., Broxton, N.M., Livett, B.G., Down, J.G., Jones, A., Alewood, P.F., Lewis, R.J., 1998. Alpha-conotoxin EpI, a novel sulfated peptide from *Conus episcopatus* that selectively targets neuronal nicotinic acetylcholine receptors. J. Biol. Chem. 273, 15667–15674.

Lu, B.S., Yu, F., Zhao, D., Huang, P.T., Huang, C.F., 1999. Conopeptides from *Conus striatus* and *Conus textile* by cDNA cloning. Peptides 20, 1139–1144.

Luo, S., Kulak, J.M., Cartier, G.E., Jacobsen, R.B., Yoshikami, D., Olivera, B.M., McIntosh, J.M., 1998. Alpha-conotoxin AuIB selectively blocks alpha3 beta4 nicotinic acetylcholine receptors and nicotine-evoked norepinephrine release. J. Neurosci. 18, 8571–8579.

McIntosh, J.M., Yoshikami, D., Mahe, E., Nielsen, D.B., Rivier, J.E., Gray, W.R., Olivera, B.M., 1994. A nicotinic acetylcholine receptor ligand of unique specificity, alpha-conotoxin ImI. J. Biol. Chem. 269, 16733–16739.

McIntosh, J.M., Dowell, C., Watkins, M., Garrett, J.E., Yoshikami, D., Olivera, B.M., 2002. Alpha-conotoxin GIC from Conus geographus, a novel peptide antagonist of nicotinic acetylcholine receptors. J. Biol. Chem. 277, 33610–33615.

Olivera, B.M., 1997. E.E. Just Lecture, 1996. Conus venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. Mol. Biol. Cell. 8, 2101–2109.

Olivera, B.M., Cruz, L.J., Hillyard, D.R., McIntosh, J.M., Santos, A.D., 1996. Process and primers for identifying nucleic acids encoding A-lineage conotoxin peptides. US Patent, 5,589,340.

Olivera, B.M., Walker, C., Cartier, G.E., Hooper, D., Santos, A.D., Schoenfeld, R., Shetty, R., Watkins, M., Bandyopadhyay, P., Hillyard, D.R., 1999. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. Ann. NY Acad. Sci. 870, 223–237.

Sandall, D.W., Satkunanathan, N., Keays, D.A., Polidano, M.A., Liping, X., Pham, V., Down, J.G., Khalil, Z., Livett, B.G., Gayler, K.R., 2003. A novel alpha-conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves in vivo. Biochemistry 42, 6904–6911.

Santos, A.D., McIntosh, J.M., Hillyard, D.R., Cruz, L.J., Olivera, B.M., 2004. The A-superfamily of conotoxins: structural and functional divergence. J. Biol. Chem. 279, 17596–17606.

Sollod, B.L., Wilson, D., Zhaxybayeva, O., Gogarten, J.P., Drinkwater, R., King, G.F., 2005. Were arachnids the first to use combinatorial peptide libraries? Peptides 26, 131–139.

Terlau, H., Olivera, B.M., 2004. Conus venoms: a rich source of novel ion channel-targeted peptides. Physiol Rev. 84, 41–68.

Wang, C.G., Cai, Z., Lu, W., Wu, J., Xu, Y., Shi, Y., Chi, C.W., 2005. A novel short-chain peptide BmKX from the Chinese scorpion *Buthus martensi* Karsch, sequencing, gene cloning and structure determination. Toxicon 45, 309–319.

Wang, C.Z., Jiang, H., Ou, Z.L., Chen, J.S., Chi, C.W., 2003. cDNA cloning of two A-superfamily conotoxins from *Conus striatus*. Toxicon 42, 613–619.

Xu, C.Q., He, L.L., Brone, B., Martin-Eauclaire, M.F., Van Kerkhove, E., Zhou, Z., Chi, C.W., 2004. A novel scorpion toxin blocking small conductance Ca^{2+} activated K^{+} channel. Toxicon 43, 961–971.