THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 17, Issue of April 23, pp. 17596–17606, 2004

Printed in U.S.A.

The A-superfamily of Conotoxins

STRUCTURAL AND FUNCTIONAL DIVERGENCE*

Received for publication, September 2, 2003, and in revised form, December 26, 2003
Published, JBC Papers in Press, December 30, 2003, DOI 10.1074/jbc.M309654200

Ameurfina D. Santos‡§, J. Michael McIntosh¶‖, David R. Hillyard, Lourdes J. Cruz‡ ‡‡, and Baldomero M. Olivera‡§§**

*From the Departments of ‡Biology, ¶Psychiatry, and **Pathology, University of Utah, Salt Lake City, Utah 84112 and the §National Institute of Molecular Biology and Biotechnology and the ‡‡Marine Science Institute, University of the Philippines, Diliman, Quezon City 1101, Philippines*

The generation of functional novelty in proteins encoded by a gene superfamily is seldom well documented. In this report, we define the A-conotoxin superfamily, which is widely expressed in venoms of the predatory cone snails (*Conus*), and show how gene products that diverge considerably in structure and function have arisen within the same superfamily. A cDNA clone encoding α-conotoxin GI, the first conotoxin characterized, provided initial data that identified the A-superfamily. Conotoxin precursors in the A-superfamily were identified from six *Conus* species: most (11/16) encoded α-conotoxins, but some (5/16) belong to a family of excitatory peptides, the κA-conotoxins that target voltage-gated ion channels. α-Conotoxins are two-disulfide-bridged nicotinic antagonists, 13–19 amino acids in length; κA-conotoxins are larger (31–36 amino acids) with three disulfide bridges. Purification and biochemical characterization of one peptide, κA-conotoxin MIVA is reported; five of the other predicted conotoxins were previously venom-purified. A comparative analysis of conotoxins purified from venom, and their precursors reveal novel post-translational processing, as well as mutational events leading to polymorphism. Patterns of sequence divergence and Cys codon usage define the major superfamily branches and suggest how these separate branches arose.

The predatory cone snails use their biochemically and pharmacologically complex venoms to capture prey, defend against predators, and compete with other animals in their environment (1). The active components of these venoms are mostly small, disulfide-rich peptides, 10–30 amino acids in length ("conotoxins") (2). As a group, cone snails (genus *Conus*) are very successful (>500 species), arguably the largest single genus of marine invertebrates (3). During speciation, a remarkable hypermutation of *Conus* venom peptides occurs (4, 5). The entire spectrum of 50–200 peptides that comprises the conotoxin repertoire of a particular *Conus* species is distinct from the peptide complement of the venom of any other *Conus* species. Thus, it is estimated that a total of >50,000 different peptides can potentially be expressed in the venoms of living *Conus*.

A simplifying conceptual framework for addressing the formidable molecular complexity of the conotoxins is that the 50,000 peptides are encoded by relatively few gene superfamilies, which have greatly diversified within this genus. In general, each conotoxin superfamily has distinctive and conserved sequence features in the conotoxin precursors encoded by the gene superfamily. Several superfamilies have been characterized, including the O-, T-, P-, I-, and M-superfamilies that have between two and five disulfide cross-links (4, 6–9).[1] The mature conotoxins in each conotoxin superfamily have a characteristic arrangement of cysteine residues, the oxidative folding of which results in a three-dimensional scaffold with a specific disulfide connectivity ("the Cys framework"). The conserved Cys superfamily framework largely determines the three-dimensional conformation of the polypeptide backbone of all peptides that belong to the superfamily.

The very first toxin purified and characterized from a *Conus* venom was α-conotoxin GI, a competitive nicotinic antagonist from the venom of the geography cone, *Conus geographus* (10). α-GI was shown to be a 13-amino acid peptide with two disulfide cross-links. The biochemical studies on α-GI established for the first time that the biologically active constituents of *Conus* venoms were small, multiply cross-linked peptides, generically referred to as conotoxins. Although several conotoxin superfamilies have been described, the gene superfamily to which α-conotoxin GI belongs has not been previously defined in the literature. We show that the A-conotoxin gene superfamily, to which α-conotoxin GI belongs, is found in fish-hunting, snail-hunting, and worm-hunting species of *Conus* and has undergone both structural and functional diversification.

EXPERIMENTAL PROCEDURES

Isolation and Sequencing of cDNA Clones of α-Conotoxin GI—cDNA clones encoding α-conotoxin GI were isolated from a size-fractionated cDNA library previously constructed by J. Hunspurger from *Conus geographus* duct mRNA using pUC13 as cloning vector and *Escherichia coli* strain MC1061F′laclq as the host cell. Clones were identified by screening them with a mixed oligonucleotide probe corresponding to the first six amino acids of GI (5′-GC*X*GG*R*TT*R*CA*R*CAYTC-3′, where *R* = A or G, *Y* = C or T, and *X* = AGT or C). The probe was end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase (11) and hybridized to nitrocellulose filters for 40–60 h at 48 °C in 3 M tetramethylammonium

* This work was supported in part by Program Project GM-48677 from the National Institutes of Health. A preliminary account of some of this work was presented in an unpublished thesis (30) as well as in the patent literature (Olivera *et al.* (12/31/96), *Process and primers for identifying nucleic acids encoding A-lineage conotoxin peptides*, U.S. Patent 5,589,340 (33)). Two of the sequences in the unpublished thesis were independently discovered by Bai-Song *et al.* (31). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ Also supported by National Institutes of Health Grant MH-53631.

§§ To whom correspondence should be addressed: Dept. of Biology, University of Utah, 257 S. 1400 E., Salt Lake City, UT 84112. Tel.: 801-581-8370; Fax: 801-585-5010; E-mail: olivera@biology.utah.edu.

[1] G. P. Corpuz, R. B. Jacobsen, E. C. Jimenez, M. Watkins, C. Walker, C. Colledge, J. E. Garrett, D. McDougal, W. Li, W. R. Gray, D. R. Hillyard, J. Rivier, J. M. McIntosh, L. J. Cruz, and B. M. Olivera, submitted for publication.

chloride (TMAC),[2] 0.1 M sodium phosphate, 1 mM EDTA, 5× Denhardt's, 0.6% SDS, and 100 μg/ml salmon sperm DNA. After washing twice in 3 M TMAC/50 mM Tris/0.2% SDS and two times in 2× standard saline citrate/0.1% SDS, positive colonies were identified by autoradiography. Some clones gave the sequence of α-GIB, a homologous α-conotoxin. α-GIB clones were eliminated by subtractive screening using an additional probe corresponding to the five amino acids at the C terminus of α-GI. Positive clones were grown in LB/amp/kan broth at 37 °C, harvested, then lysed with alkali (12). Plasmid DNA was purified by equilibrium centrifugation in CsCl ethidium bromide gradient. Sequencing of inserts in both strands was done by the dideoxy method (13) using Sequenase version 2.0 enzyme from US Biochemicals. The sequence shown in Table I was obtained by walking through the cDNA with five primers in the forward direction and five primers in the reverse direction.

Cloning and Sequencing of A-superfamily Peptides from Several Conus Species—The cDNAs were prepared from cDNA libraries of *Conus striatus* and *C. magus*, and the poly(A)$^+$ mRNA obtained from the venom ducts of *C. striatus*, *C. stercusmuscarum*, *C. bandanus*, and *C. caracteristicus*; mRNA was from the venom ducts of several snails in each case. The cDNA libraries were prepared using either the protocol in Sambrook (11) or the Promega Magic Miniprep system. cDNAs were amplified using a pair of primers, one corresponding to the signal sequence and the other to the 3′-untranslated region near the open reading frame. Amplification with Vent polymerase was done for 30 cycles using an air thermocycler set as follows: denaturation at 94 °C, 0 s; annealing at 50 °C, 0 s; elongation at 72 °C, 15 s. Ten-microliter aliquots of PCR products were analyzed by electrophoresis in 1.5% agarose gel in Tris-acetate buffer with ethidium bromide, and the remaining products were purified by gel electrophoresis in a 1.5% low melting gel.

The plasmid vector used for cloning was a modified pDH52, which we named pDH78. pDH52 is basically pGEM 3Zf+ (Promega) with its polylinker region extended by cloning into the EcoRI and KpnI sites a synthetic oligomer encoding sites for EcoRI, SacI, EcoRV, NotI, SphI, NsiI, StuI, BglII, and KpnI. This was further modified by adding a 23-mer synthetic DNA fragment that has a KpnI site inserted into the XbaI and BglII sites to generate pDH78. Plasmid DNA digested with KpnI and the PCR-amplified products were treated with 2 units of T4 DNA polymerase per milligram of DNA in the presence of 0.5 mM dATP for the vector and 0.5 mM dTTP for the PCR product in a 10- to 20-μl solution of 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 μg/μl bovine serum albumin. The mixture was incubated at 37 °C for 30 min, then heated for 10 min at 70–75 °C to stop the reaction. For ligation, a mixture of DNA polymerase-treated vector (0.1 μg) and PCR product (0.05 μg) was preincubated at 45 °C for 5 min before 5 nmol of ATP and 7–20 units of ligase were added, and the mixture was incubated at 16 °C for 16 h. Transformants were tested for the presence of the desired recombinant plasmids by digestion of the purified plasmid with XbaI and BglII, by PCR of the purified plasmid and/or by PCR of the colony after lysis. Solid phase sequencing (12) was done using magnetic beads from Dynal and the USB DNA sequencing kit.

Purification of κA-conotoxin MIVA—Freeze-dried *Conus magus* venom (500 mg) was extracted with 0.1% trifluoroacetic acid and filtered through Centriprep 30 microconcentrators as previously described by Cartier *et al.* (14). The extract was fractionated on a semi-preparative Vydac C18 column equipped with a guard module. All other chromatographic purifications involved an analytical Brownlee C_8 column (4.6 mm × 22 cm, RP300 packing, 7-mm particle size, 300-Å pore size) with a guard cartridge that also had RP300 packing material. For all chromatograms, buffer A was 0.1% trifluoroacetic acid and buffer B was 60% acetonitrile in 0.1% trifluoroacetic acid. The activity of fractions was monitored by measuring synaptically evoked responses from cutaneous pectoris muscle of frog as previously described (15–17).

Pyridylethylation, Chymotrypsin Digestion, and Sequencing of MIVA—The peptide was pyridylethylated prior to sequencing. An aliquot of HPLC fraction containing 1 nmol of MIVA was concentrated down to 5 μl of MIVA and taken up in 50 μl of 0.25 M Tris/2 mM EDTA/6 M guanidine HCl, pH 7.7, mixed with 2 μl of 10% β-mercaptoethanol, and incubated at room temperature for 30 min. Two microliters of 20% 4-vinylpyridine in ethanol was added, and the mixture was incubated for 15 min at room temperature in the dark. 25 μl of 2% heptafluorobutyric acid (HFBA) was added prior to HPLC on an analytical Brownlee C_8 column (RP300, 220 × 4.6 mm), which was eluted with a gradient of acetonitrile in 0.05% HFBA.

FIG. 1. ***Conus* species that were analyzed for A-superfamily conotoxins.** Shells of four fish-hunting species are shown: *top row, left* to *right*, *C. striatus*, *C. geographus*, and *C. stercusmuscarum*; *bottom row, middle*, *C. magus*. One mollusc-hunting species is shown in the *bottom row, left*, *C. marmoreus*, and one worm-hunting species is at the *bottom row right*, *C. caracteristicus*.

The pyridylethylated peptide, concentrated to ~5 μl, was mixed with 2.5 μl of acetonitrile before addition of the digestion buffer (0.25 M Tris/10 mM $CaCl_2$, pH 7.8) and 2 μl of 0.05 μg/μl chymotrypsin (Roche Applied Science). The reaction mix was incubated overnight at room temperature, 25 μl of 2% HFBA was added before separation of the fragments by HPLC as described above for the pyridylethylated peptide. Sequencing of the alkylated peptide and the chymotrypsin fragments was performed on an Applied Biosystems model 477A protein sequencer at the Protein/DNA Core Facility at the University of Utah Cancer Center.

RESULTS

Identification and Analysis of a cDNA Clone Encoding α-Conotoxin GI—A cDNA sequence of the α-conotoxin GI prepropeptide matched the canonical organization originally established for conotoxin gene superfamilies by Hillyard and coworkers (4); the literature on conotoxin superfamilies was recently reviewed (1). The messenger RNA encoding α-conotoxin GI has the characteristic four-module organization, including a signal sequence, a "pro" region, the toxin-encoding region followed by a stop codon, and a 3′-untranslated region. For the α-conotoxin GI cDNA, the signal sequence is 21 amino acids, the pro region is 28 aa and the C-terminal toxin region comprises 15 amino acids that are predicted to be post-translationally processed to the 13-amino acid toxin. For α-conotoxin GI, the 3′-untranslated region is quite long: 823 nucleotides (Fig. 1). The generation of the mature toxin requires proteolytic cleavage of the N-terminal prepro region of the precursor, and standard processing at the C terminus from the sequence -CGR- to -C-NH_2.

[2] The abbreviations used are: TMAC, tetramethylammonium chloride; α-, α-conotoxin; HFBA, heptafluorobutyric acid; HPLC, high-performance liquid chromatography; aa, amino acid(s).

TABLE I

cDNA sequence of the gene for α-conotoxin GI and the predicted translation product

The signal sequence is shaded, and the toxin region is underlined. Nucleotides in the untranslated region are in small letters, and the poly(A) addition signal (aataaa) is shaded. In α-GIB, two amino acids (KG) corresponding to an insertion of six nucleotides (AAG GGA) are in the position marked by a dash (—) in the above sequence of α-GI.

```
acc ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTG GTC TTG GCA ACC ACT GTC GTT TCC TTC
    M   G   M   R   M   M   F   T   V   F   L   L   V   V   L   A   T   T   V   V   S   F

CCT TCA GAA CGT GCA TCT GAT GGC AGG GAT GAC ACA GCC AAA GAC GAA GGG TCT GAC ATG GAG AAA TTG
 P   S   E   R   A   S   D   G   R     D   D   T   A   K   D   E   G   S   D   M   E   K   L

GTC GAG AAA AAA GAA TGT TGC AAT CCT GCC TGT GGC AGA CAC TAC AGT TGT — GGA CGC TGA
 V   E   K   K   E   C   C   N   P   A   C   G   R   H   Y   S   C   —   G   R  STOP

tgctccagga ccctctgaac cacgacgtgc cgccctctgc ctgacctgct tcactgtccg tctctttgtg ccactagaac

tgaacagctc gatccactag actaccacgt tacctccgtg ttctaaaact acttggttta gattgccttt aatttctagt

catacttcct gttattacgt cgtccaaaat tgaaacaaga acatgagggg tgtcagctca aacaaaatca ggcaatgaca

aggaaaatgt ctccgatcga tccgaaaact gtcacccgtc actctcttaa ccagttttag aactgattac cactagagct

tttgtaccac atcaaatcag gtctatgtgt gatgtttctt ttgcaaaatt taatttttga gaaaaaaagc tcaaaatgtg

ggaagtgctt ttgatttttct gacaacttgt gatcatgtcc gttttcagtg agtctaattg caacctctgt gtgatttttct

tcacctgtta agcaacgcaa agaggttgtc cataaccagg aaagcaacag acaaagaaat gcttgagaat ttcaggttat

agataaggta aggaaaaaaa ggagagctat gggaaatgat gaaaacaaca gataaaataa attgaacagt acctacttgt

ttcatggttg attttttttt ctctgaataa tctctgtgga cactaatggc agtctctcct caccccacgc cattagtaag

cttatttttt ctttctttat ccaagatttg ctgaacatat ttagcctaga tatagacatt gctacatata taatctgaca

ataaactttc atgggcacca attpoly(A)
```

A closely related cDNA sequence was also found, which appears to be a polymorphism in the α-conotoxin GI gene. The nucleotide sequence is identical except for an insertion of six nucleotides, as shown in Table I, where the relevant α-GI sequence and the cDNA encoding the variant (designated here as α-conotoxin GIB) are shown. The insertion of the six nucleotides would lead to a mature toxin predicted to have two extra amino acids at the C terminus. Given the sequence identity between them, α-GI and α-GIB may be allelic variants.

Characterization of Related cDNA Clones from Another Fish-hunting Species, C. striatus—To compare related genes from another Conus species, we characterized cDNA clones from the venom duct of the striated cone, *C. striatus*, a fish-hunting species (see Fig. 1). This venom has been extensively characterized (18), and several α-conotoxins were previously purified from the venom of this species.

cDNA clones were analyzed from a *C. striatus* venom duct cDNA library as described under "Experimental Procedures." Five distinct sequences were obtained, which are compared with α-conotoxin GI in Table II. As expected, two of these sequences encode α-conotoxins previously characterized from *C. striatus* venom. These are α-conotoxin SI and α-conotoxin SII. The latter is an unusual peptide closely related in sequence to other α-conotoxins, but with an additional disulfide bond. A comparison of the amino acid sequences of the three predicted open reading frames (for α-conotoxins GI, SI, and SII) reveals that the signal sequence regions are identical. In general, sequence divergence increases from the N to C terminus of the open reading frame; in the signal sequence, there is at most a single amino acid that differs from the 21 aa consensus with greater divergence in the pro region, and considerable hypervariability at the C-terminal toxin-encoding region. There are clearly three groups of toxin sequences based on sequence similarity: the first four sequences (α-conotoxins GI, GIB, SI, and SII) show much greater sequence homology to each other than to the other two groups in Table II (S1.1 and SIVA/SIVB). The degree of sequence divergence is reflected in the biological properties of the peptides (see "Discussion").

In addition to the α-conotoxin SI- and SII-encoding cDNA clones, the S1.1 clone from *C. striatus* encoded the precursor of an α-conotoxin-like peptide. Although the predicted mature peptide has the characteristic Cys pattern of the α-conotoxins (–CC–C–C–), it is very divergent in amino acid sequence from α-conotoxins GI, SI, and SII (which share considerable sequence homology). The predicted mature peptide has not been purified from venom.

A surprise from the analysis of the *C. striatus* cDNA was a class of clones that exhibited substantial sequence identity to those above but did not encode α-conotoxin precursors; instead, these encode two κA-conotoxins. One of the predicted mature conotoxins, κA-conotoxin SIVA, was previously characterized from venom (17). The two such sequences obtained originated from two different collection samples: the clones encoding κA-SIVA were obtained from a cDNA library made from snails collected in the Philippines. A second set of closely related clones, encoding a peptide designated as κA-SIVB was obtained from cDNA made from specimens of *C. striatus* collected in the Hawaiian Islands. As shown in Table II, the N-terminal sequences of these clones have considerable sequence identity to the clones that encode the α-conotoxins: the seven signal sequences (α-conotoxins GI, GIB, SI, and SII and κA-conotoxins SIVA, SIVB, and S1.1) were found to be virtually identical (see

TABLE II

Comparison of conotoxin prepropeptide sequences

Signal sequences are shaded, and the toxin regions are underlined. In A, the shaded sections of the cDNA sequences mark the 25-nucleotide alteration that explains the difference in length of κA-SIVA and κA-SIVB, and the 6-nucleotide change at the carboxyl terminus of SI to generate a longer conotoxin (SII). The sequence of α-conotoxin GI from *C. geographus* is given for comparison.

A. Prepropeptide sequences of α- and κA-conotoxins from *Conus striatus*, a fish-hunting species

α-conotoxins (α3/5)

```
GI      MGMRMMFTVFLLVVLATTVVSFPSERASDGRDDTAKDEGSDM-EKLVEKKE-CCNPACGRHYSCGR
GIB     MGMRMMFTVFLLVVLATTVVSFPSERASDGRDDTAKDEGSDM-EKLVEKKE-CCNPACGRHYSCKGGR
SI      MGMRMMFTVFLLVVLATTVVSFPSDRASDGRDDEAKDERSDMHESD--RKEICCNPACGPKYSCGR
SII     MGMRMMFTVFLLVVLATTVVSFPSDRASDGRDDEAKDERSDMHESDRNGRGCCCNPACGPNYGCGTSCSRTL
```

α-conotoxin (α4/4)

```
SI.1    MGMRMMFTVFLLVVLAITVVSFPLDRESDGANAEARTHDHEKHALDRNGCCRNPACESHRCG
```

κA-conotoxins

```
SIVA    MGMRMMFTVFLLVVLATNVVSTPSDRASDGRNAAVHERQKSLVPSVITTCCGYDPGTMCPPCRCTNSCG
SIVB    MGMRMMFTVFLSVVLATTVVSTPSDRASDGRNAAVHERQKELVPSVITTCCGYDPGTMCPPCRCTNSCPTKPKKPGRRND
```

cDNA sequences of the C-terminal regions

```
SIVA    5'--- TGC ACT AAT AGC TGT GGT TAA cca aaa ccc aaa aca ggc cgc aga aac gac tga ---3'
              C   T   N   S   C   G  STOP

SIVB    5'--- TGC ACT AAT TCC TGT CCA ACA AAA CCG AAA AAA CCA GGC CGC AGA AAC GAC TGA ---3'
              C   T   N   S   C   P   T   K   P   K   K   P   G   R   R   N   D  STOP

SI      5'--- GCC TGT GGC CCA AAG TAT AGT TGT GGA CGC TGA tgc tcc agg acc ctc tga acc ---3'
              A   C   G   P   K   Y   S   C   G   R  STOP

SII     5'--- GCC TGT GGC CCA AAC TAT GGT TGT GGC AGC TCA TGC TCC AGG ACC CTC TGA acc ---3'
              A   C   G   P   N   Y   G   C   G   T   S   C   S   R   T   L  STOP
```

B. Prepropeptide sequences of conotoxins from the fish-hunting species *Conus magus* and *Conus stercusmuscarum*

α4/7 prepropeptide

```
MII     MGMRMMFTVFLLVVLATTVVSFPSDRASDGRNAAANDKASDVITLALKGCCSNPVCHLEHSNLCGRRR
```

κ-conotoxin prepropeptides

```
MIVA    MGMRMMFTVFLLVVLATTVVSIPSDRASDGRNAVVHERAPELVVTATTNCCGYNPMTICPPCMCTYSCPPKRKPGRRND
SmIVA   MGMRMMFTVFLLVVLATTVVSIPSDRASDGRNAAVNERQTWLVPSTITTCCGYDPGTMCPTCMCDNTCKPKPKKSGRRND
SmIVB   MGMRMMFTVFLLVVLATTVVSIPSDRASDGRNAEVNERAPWLVPSTITTCCGYDPGSMCPPCMCNNTCKPKPKKSGRRNH
```

C. Prepropeptide sequences of conotoxins from the non-fish-hunting species *Conus bandanus* and *Conus caracteristicus*

```
Bn1.1     MGMRMMFTMFLLVVLATTVVSFASDRASDGRNAAAKDKASDLVALTVKGCCSHPACSVNNPDICG
Bn1.2                                                  ...KECCTHPACHVSHPELCG
Bn1.3     MGMRMMFTVFLLVVLATAVLPVTLDRASDGRNAAANAKTPRLIAPFIRDYCCHRGPCMVWCG
Ca1.1     MGMRMMFTVFLLVVLATTVVSFTSDRASDGRNAAANAFD--LIALIARQNCCSIPSCWEKYKCS
Ca1.2     MGMRMMFTVFLLVVLATTVVSFTSDRASEGRNAAAKDKASDLVALTVRGCCAIRECRLQNAAYCGGIY
```

shaded amino acid sequences in Table IIA). However, in the propeptide region there is more divergence between the clones that encode α-conotoxins *versus* those that encode κA-conotoxins (see "Discussion"), and, in the encoded mature toxins (*underlined sequences*), there is essentially no homology detectable between the α-conotoxins and κA-conotoxins.

The two clones encoding κA-conotoxins are largely identical in the predicted amino acid sequence for the first 68 amino acids of the open reading frame (only 2 aa substitutions). However, the mature toxin region of κA-SIVA comprises 31 aa, whereas for κA-SIVB it is 42 aa. Thus, the SIVA precursor is a total of 69 aa *versus* the SIVB precursor, which is 80 aa. The C-terminal amino acid of SIVA is predicted to be a glycine followed by a stop codon. As shown in Table IIA, there is no stop codon at the homologous position in SIVB, resulting in an additional 11 amino acids at the C-terminal end. An examina-

tion of the nucleotide sequence in this region, shown in Table II, indicates that the conversion from SIVA to SIVB is not explained by a simple substitution, insertion, or deletion; instead, there is a substitution of a block of 25 nucleotides (*shaded nucleotide sequences* in Table II*A*). At the 3′-end of the open reading frame of SIVB, the sequences become identical once again.

The discovery of the considerable extent of sequence identity between the α-conotoxin and κA-conotoxin precursors is unprecedented. The arrangement of Cys residues in the mature toxins are completely different for the two groups; furthermore, the κA-conotoxins are excitatory peptides targeted to voltage-gated ion channels and not nicotinic receptor antagonists, the physiological mechanism underlying the activity of α-conotoxins (19). Thus, although the two groups of peptides are genetically related (with considerable identity in signal sequences and the N-terminal side of their propeptide regions) and belong to the same gene superfamily, they have clearly diverged to yield two families of conotoxins that are both structurally and functionally distinct.

Analysis of cDNA Clones from Other Fish-hunting Conus Species—To explore which members of the A-conotoxin superfamily are highly expressed in other fish-hunting species, we investigated *C. magus*, the magician's cone, and *Conus stercusmuscarum*, the fly-speck cone (see Fig. 1). The results of the analysis are shown in Table II*B*. From *C. magus*, one open reading frame encoding a precursor sequence for an α-conotoxin was identified; surprisingly, all of these encoded an α-conotoxin MII precursor. α-Conotoxin MII has been extensively characterized (14, 20) and shown not to interact with the nicotinic receptor at the neuromuscular junction. Instead, α-MII potently blocks two neuronal nicotinic receptor subtypes, α3β2 and α6β3β2 (14, 21, 22).

The second A-superfamily sequence from *C. magus* encoded a peptide that had high sequence homology to κA-conotoxin SIVA, which we are designating κA-conotoxin MIVA. As will be described below, we have directly characterized κA-conotoxin MIVA from *C. magus* venom. The *C. stercusmuscarum* cDNA library did not yield any α-conotoxin precursors; all of the clones analyzed had open reading frames that encoded precursors of peptides homologous to κA-conotoxins SIVA and MIVA (designated κA-conotoxins SmIVA and SmIVB). These share considerable sequence identity and are likely to be allelic variants. The predicted open reading frames are shown in Table II*B*.

Analysis of the A-superfamily in Non-piscivorous Conus Species—To compare the A-superfamily in *Conus* species that do not prey on fish to the A-superfamily conotoxins expressed in the fish-hunting *Conus* species investigated above, we identified A-superfamily members from cDNA libraries from a molluscivorous (snail-hunting) species, *Conus bandanus*, and from a vermivorous (worm-hunting) species, *Conus caracteristicus* (see Fig. 1). Using the nomenclature of Espiritu *et al.* (23), all of the fish-hunting species analyzed belong to either Clades I or II, whereas *C. bandanus* belongs to the mollusc-hunting Clade VI, and *C. caracteristicus* belongs to the worm-hunting Clade XIV. The analysis of cDNA clones yielded three different types of sequences for *C. bandanus* and two for *C. caracteristicus* (see Table II*C*).

It is noteworthy that all five sequences elucidated from these species are predicted to encode α-conotoxins; however, none of the clones analyzed encoded a three-disulfide-cross-linked-containing *Conus* peptide (such as κA-conotoxin SIVA). Thus, in contrast to such species as *C. striatus*, *C. magus*, and *C. stercusmuscarum*, the non-piscivorous species we have analyzed are notable in encoding only α-conotoxin-like sequences.

None of the *C. caracteristicus* or *C. bandanus* peptides have been purified from venom. However, one of the *C. bandanus* peptides, encoded by clone Bn1.3, was chemically synthesized as described under "Experimental Procedures" and tested on six different cloned rat nicotinic acetylcholine receptor subtypes expressed in *Xenopus* oocytes by previously described methods (14) (α2β2, α2β4, α3β4, α4β2, α4β4, and α6/α3β2β3). The peptide did not detectably affect any of the nicotinic receptor subtypes tested ($IC_{50} > 5$ μM for all subtypes). The lack of activity may be due to high specificity of the peptide for a molluscan target, although the possibility that a post-translational modification in the native peptide is required for activity is an alternative explanation (see next section).

Isolation and Characterization of κA-conotoxin MIVA—A highly potent excitatory peptide that elicits repetitive action potentials and causes the same spastic symptomatology as κA-conotoxin SIVA was purified from *C. magus* venom (Fig. 2). As shown in Table III, initial sequencing of the pyridylethylated peptide gave a 36-amino acid peptide with a faint Glu signal in position 3 and blanks in positions 7 and 9. The sequence is otherwise identical to that predicted from a cDNA clone from *C. magus* (*underlined sequence* in Table II*B*), indicating that positions 7 and 9 are modified Thr residues.

Chymotrypsin digestion of the peptide gave three main fragments. Fragments 1 and 1a are essentially the same except for position 3, where a clear signal for Glu was found in fragment 1 and only a trace of Glu was found in fragment 1a, which is routinely observed when γ-carboxyglutamate (Gla) is present. This is consistent with the mass spectrometry data that indicated a mass difference of 44 between fragments 1 and 1a. In fragment #2, there was an apparent cleavage after the Met in addition to the cleavage after the Tyr. Fragment #3 clearly ends with Hyp, suggesting that the GRR serves as an amidation signal. Amidation of this fragment was confirmed by mass spectrometry. The mass of the peptide obtained by liquid secondary ionization mass spectrometry (average $MH^+ = 5095.3$) was 1053.6 Da higher than the mass predicted for the peptide containing Gla in position 3 and an amide group at the carboxyl end. This mass difference was localized to fragment 1 where the two threonines predicted from cDNA clones appear as blanks upon chemical sequencing. This is consistent with the two blank positions (7 and 9) being glycosylated Thr. The "missing mass" of 1053.6 presumably can be divided between two glycosylation sites on Thr^7 and Thr^9. The combination of nucleotide sequencing, amino acid sequencing, and mass spectrometry data confirms the sequence of κA-conotoxin MIVA as AOγLVVTATTNCCGYNOMTICOOCMCTYSCOOKRKO-NH_2 (γ, γ-carboxyglutamate; T, modified threonine-putative *O*-glycosylation; O, hydroxyproline).

DISCUSSION

The small pharmacologically active peptides of *Conus* venoms have shown remarkable interspecific divergence. The ~50,000 different peptides found in *Conus* venoms can be grouped into *Conus* peptide families; in turn, several different *Conus* families may belong to a *Conus* peptide gene superfamily. Each *Conus* peptide gene superfamily has a number of shared diagnostic elements, the most striking of which is the highly conserved signal sequence. In contrast to families of conventional secreted gene products, in which the signal sequences at the N terminus are the least conserved elements, in conotoxin gene superfamilies the reverse is true: signal sequences are the most highly conserved region. Thus, the signal sequence is diagnostic of which superfamily a particular *Conus* peptide belongs to.

The A-conotoxin superfamily is defined here by 16 sequences from four species of fish-hunting *Conus* (*C. geographus*,

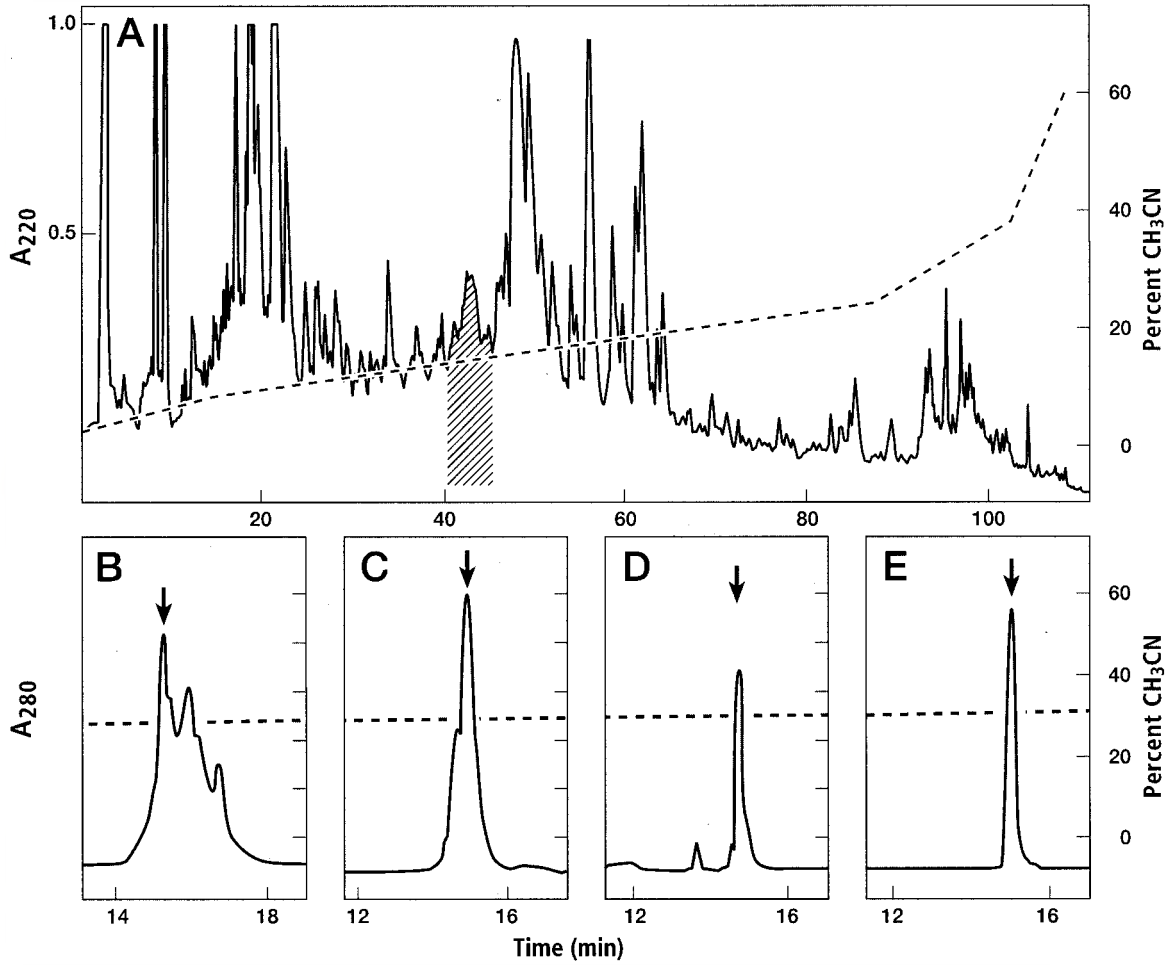

FIG. 2. **Purification of κA-Conotoxin MIVA.** Samples were fractionated on C_{18} HPLC columns eluted with a gradient of acetonitrile in 0.1% trifluoroacetic acid indicated by the *dashed line*. *A*, fractionation of the venom extracts on semi-preparative C_{18} HPLC. *B*, fractions corresponding to the *shaded area* in *A* were pooled and applied on a C_{18} analytical column. *C*, fractions from seven HPLC runs corresponding to the peak marked by the *arrow* in *B* were pooled and chromatographed on the C_{18} analytical column. *D*, the *peaks* (marked by an *arrow* in *C*) from five runs were pooled and rerun on the C_{18} column. *E*, the peak marked by the *arrow* in *D* was cleaned by another run on the analytical column. Absorbance was monitored at 220 nm in *A* and at 280 nm in *B–E*. The active peaks (marked by *arrow*) eluted at 14.7–14.8 min in 30% acetonitrile-0.1% trifluoroacetic acid.

TABLE III
Sequencing and mass spectrometry of κ-A-conotoxin MIVA

		Sequence	Average mass		
			Predicted	Observed	Difference
					Da
Sequence predicted from cDNA clone[a]		APELVVTATTNCCGYNPMTICPPCMCTYSCPPKRKP-NH_2	4041.7	5095.3	1053.6
Sequence of alkylated peptide		AOγLVV-A-TNCCGYNOMTICOOCMCTYSCOOKRKO			
Chymotrypsin fragment	#1	AOELVV-A-TNCCGY	1769.0[b]	2822.3	1053.3
	#1a	AO-LVV-A-TNCCGY	1813.1[b]	2866.0	1053.0
	#2	TICOOCMCTY	1479.8	1479.2	
	#3	SCOOKRKO	1064.6	1064.6	
Sequence of κA-conotoxin MIVA[c]		AOγLVVTATTNCCGYNOMTICOOCMCTYSCOOKRKO-NH_2			

[a] The predicted sequence assumes that PGRRND is post-translationally modified to P-NH_2.
[b] Estimated assuming that residue 3 is γ-carboxyglutamate in fragment #1a, and residues 7 and 9 are threonines.
[c] The putative glycosylated threonine residues are underlined. *O* is hydroxyproline; γ is γcarboxyglutamate.

C. striatus, *C. magus*, and *C. stercusmuscarum*) as well as from two non-fish-hunting species (*C. bandanus*, a snail-hunting *Conus*, and *C. caracteristicus*, a worm-hunting *Conus*).[1] All species examined yielded cDNA clones that encode A-conotoxin superfamily precursors; these precursors have the diagnostic N-terminal signal sequence, an intervening pro region and a C-terminal mature toxin region. The exact 21-amino acid consensus sequence for the N-terminal signal sequence (MGMRMMFTVFLLVVLATTVVS) is found without any substitution whatsoever in a large majority of the different precursor sequences analyzed. This highly conserved signal sequence serves as a defining feature for members of the A-conotoxin gene superfamily. In striking contrast, there is a dramatic divergence of mature conotoxin sequences. To give a specific example, α-conotoxin MII and κA-conotoxin MIVA both from *C. magus* have identical signal sequences, but there is little, if any, homology in the mature toxin sequences (Table IV).

Also shown in Table IV is a comparison of the nucleotide sequences of the signal sequence region, contrasted with the nucleotide sequences of mature toxin regions. It is noteworthy

TABLE IV

Comparison of nucleotide and amino acid sequences of signal sequence and mature toxin regions

In A: Z, pyroglutamate; O, hydroxyproline; γ, gamma-carboxyglutamate; underlined amino acids are glycosylated. All post-translational modifications are assumed to be homologous to isolated κA-conotoxins. Sequence differences are shaded in A.

A. Amino acid sequences

		Predicted or Actual* Mature Toxin	Signal Sequence
α3/5			
	GI*	ECCNPACGRHYSC-NH_2	MGMRMMFTVFLLVVLATTVVS
	GIB	ECCNPACGRHYSCKG-NH_2	MGMRMMFTVFLLVVLATTVVS
	SI*	EICCNPACGPKYSC-NH_2	MGMRMMFTVFLLVVLATTVVS
	SII*	GCCCNPACGPNYGCGTSCSRTL	MGMRMMFTVFLLVVLATTVVS
α4/7			
	MII*	GCCSNPVCHLEHSNLC-NH_2	MGMRMMFTVFLLVVLATTVVS
	Bn1.1	GCCSHPACSVNNPDIC-NH_2	MGMRMMFTMFLLVVLATTVVS
	Bn1.2	ECCTHPACHVSHPELC-NH_2	
	Ca1.2	GCCAIRECRLQNAAYCGGIY	MGMRMMFTVFLLVVLATTVVS
Other α-conotoxins			
	Bn1.3 (α4/3)	DYCCHRGPCMVWC-NH_2	MGMRMMFTVFLLVVLATAVLP
	S1.1 (α4/4)	NGCCRNPACESHRC-NH_2	MGMRMMFTVFLLVVLAITVVS
	Ca1.1 (α4/5)	ZNCCSIPSCWEKYKCS	MGMRMMFTVFLLVVLATTVVS
κA			
	MIVA*	AOγLVVTATTNCCGYNOMTICOOCMCTYSCOOKRKO-NH_2	MGMRMMFTVFLLVVLATTVVS
	SIVA*	ZKSLVPSVITTCCGYDOGTMCOOCRCTNSC-NH_2	MGMRMMFTVFLLVVLATNVVS
	SIVB	ZKγLVPSVITTCCGYDOGTMCOOCRCTNSCOTKOKKO-NH_2	MGMRMMFTVFLSVVLATTVVS
	SmIVA	ZTWLVPSTITTCCGYDOGTMCOTCMCDNTCKOKOKKS-NH_2	MGMRMMFTVFLLVVLATTVVS
	SmIVB	AOWLVPSTITTCCGYDOGSMCOOCMCNNTCKOKOKKS-NH_2	MGMRMMFTVFLLVVLATTVVS

B. Nucleotide sequences

Signal sequence

α3/5

GI/GIB	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTG GTC TTG GCA ACC ACT GTC GTT TCC
SI/SII	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTT GTC TTG GCA ACC ACT GTC GTT TCC

α4/7

MII	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTT GTC TTG GCA ACC ACT GTC GTT TCC
Bn1.1	ATG GGC ATG CGG ATG ATG TTC ACC ATG TTT CTG TTG GTT GTC TTG GCA ACC ACT GTC GTT TCC
Ca1.2	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTT GTC TTG GCA ACC ACT GTG GTT TCC

κA

MIVA/SmIVA	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTT GTC TTG GCA ACC ACT GTC GTT TCC
SIVA	ATG GGC ATG CGG ATG ATG TTC ACC GTG TTT CTG TTG GTT GTC TTG GCA ACC AAT GTC GTT TCC

Toxin region

α3/5

GI	GAA **TGT TGC** AAT CCT GCC **TGT** GGC AGA CAC TAC AGT **TGT** GGA CGC TGA
GIB	GAA **TGT TGC** AAT CCT GCC **TGT** GGC AGA CAC TAC AGT **TGT** AAG GGA GGA CGC TGA
SI	GAA ATC **TGT TGC** AAT CCT GCC **TGT** GGC CCA AAG TAT AGT **TGT** GGA CGC TGA
SII	GGA **TGC TGT TGC** AAT CCT GCC **TGT** GGC CCA AAC TAT GGT **TGT** GGC ACC TCA **TGC** TCC AGG ACC CTC TGA

α4/7

MII	GGA **TGC TGT** TCC AAC CCT GTC **TGT** CAC TTG GAG CAT TCA AAC CTT **TGT** GGT AGA AGA CGC TGA
Bn1.1	GGA **TGC TGT** TCT CAT CCT GCC **TGT** AGC GTG AAT AAT CCA GAC ATT **TGT** GGT TGA
Bn1.2	GAA **TGC TGT** ACT CAT CCT GCC **TGT** CAC GTG AGT CAT CCA GAA CTC **TGT** GGT TGA
Ca1.2	GGA **TGC TGT** GCC ATT CGT GAA **TGT** CGC TTG CAG AAT GCA GCG TAT **TGT** GGT GGA ATA TAC TGA

TABLE IV—*continued*

Nucleotide sequences	
Other α-conotoxins	
Bn1.3	GAT TAT **TGC TGT** CAT AGA GGT CCC **TGT** ATG GTA TGG **TGT** GGT TGA
S1.1	AAT GGA **TGC TGT** AGG AAT CCT GCC **TGT** GAG AGC CAC AGA **TGT** GGT TGA
Ca1.1	CAA AAT **TGC TGT** AGC ATT CCC AGC **TGT** TGG GAG AAA TAT AAA **TGT** AGT TAA
κA	
MIVA	GCG CCT GAG CTG GTC GTT ACG GCC ACC ACG AAT **TGC TGT** GGT TAT AAT CCG ATG ACA ATA **TGC** CCT
SIVA	CAG AAG AGT CTG GTC CCT TCG GTC ATC ACG ACT **TGC TGT** GGA TAT GAT CCG GGG ACA ATG **TGC** CCT
SIVB	CAG AAG GAG CTG GTC CCT TCG GTC ATC ACG ACT **TGC TGT** GGA TAT GAT CCG GGG ACA ATG **TGC** CCT
SmIVA	CAA ACT TGG CTG GTC CCT TCG ACA ATC ACG ACT **TGC TGT** GGA TAT GAT CCG GGG ACA ATG **TGC** CCT
SmIVB	GCG CCT TGG CTG GTC CCT TCG ACA ATC ACG ACT **TGC TGT** GGA TAT GAT CCG GGG TCA ATG **TGC** CCT
MIVA	CCT **TGC** ATG **TGC** ACT TAT TCC **TGT** CCA CCA AAA AGA AAA CCA GGC CGC AGA AAC GAC TGA
SIVA	CCT **TGC** AGG **TGC** ACT AAT AGC **TGT** GGT TAA
SIVB	CCT **TGC** AGG **TGC** ACT AAT TCC **TGT** CCA ACA AAA CCG AAA AAA CCA GGC CGC AGA AAC GAC TGA
SmIVA	ACT **TGC** ATG **TGC** GAT AAT ACC **TGT** AAA CCA AAA CCC AAA AAA TCA GGC CGC AGA AAC GAC TGA
SmIVB	CCT **TGC** ATG **TGC** AAT AAT ACC **TGT** AAA CCA AAA CCC AAA AAA TCA GGC CGC AGA AAC CAC TGA

that not only are the amino acids of the signal sequences conserved, but codon usage at every locus is conserved. Because some of the sequences being compared are from pairs of *Conus* species believed to have diverged in the Miocene period or earlier (1, 23), the lack of even silent substitutions is notable.

The members of the O-conotoxin superfamily have a similarly highly conserved signal sequence; however, the O-superfamily signal sequences show no homology whatsoever to the A-superfamily signal sequences (4). Thus, each *Conus* peptide superfamily has its own characteristic signature signal sequence. A second diagnostic feature of members of a particular superfamily is the arrangement of cysteine residues, which in turn predicts the disulfide-bonding framework of the peptides belonging to that superfamily. Although most loci in the mature toxin region are hypervariable, the cysteine residues are conspicuously conserved; in general, at each Cys locus only one of the two codons is used, a feature previously noted for other conotoxins (24). Apparent exceptions are α-GI and α-SI; the first two Cys residues have an apparent reversal in codon usage from all of the other A-superfamily conotoxins in this study (TGT TGC in α-conotoxins GI and SI *versus* TGC TGT in all others; see Table IV). The possible origins of this codon reversal are discussed below. It should also be noted that cysteines 3, 4, and 5 in the κA conotoxins use TGC, whereas except for the first Cys residue, all other Cys residues are encoded by TGT in most other A-superfamily peptides.

Post-translational Processing—All conotoxins require processing from their precursors by proteolytic cleavage at a characteristic sequence (usually two basic amino acids, or alternatively –*X*R–) to yield the N terminus of the mature conotoxin. At the C-terminal end, sequences such as –*X*G, –*X*GR, –*X*GK, –*X*GRR, etc. are processed to –*X*-NH_2 by a well characterized enzyme system (25, 26). The data presented here suggest that processing at both ends can be more complex for some members of the A-conotoxin superfamily. Some cases where one might not have predicted *a priori* from the cDNA clone what mature conotoxin sequence would be found in venom are shown in Table V. Thus, for α-conotoxin SI, there is unusual processing at the N-terminal end. For α-conotoxin SII, endoproteolytic cleavage after the arginine, followed by a carboxypeptidase removal of the C-terminal Arg residue presumably occurred.

The most complex case of post-translational processing is observed in κA-conotoxin MIVA, as indicated in Table V. The proteolytic cleavage at the N terminus is conventional, but at the C terminus, initial cleavage after an arginine residue followed by the removal of an Arg by a carboxypeptidase and processing of the transiently generated Cys-Gly terminus to Cys-NH_2 presumably takes place. In addition, in contrast to α-conotoxins in which only standard amino acids are found, κA-conotoxin MIVA has ten post-translationally modified amino acids. These include seven Pro residues that have been hydroxylated, a Glu residue at position 3 that has been γ-carboxylated to Gla, and two modified residues of Thr at positions 7 and 9 that may be *O*-glycosylated.

In two peptides of this family directly purified from venom, κA-SIVA and a protein from *Conus consors* (27), a single Pro residue is not hydroxylated; this is immediately adjacent to a putative glycosylated Ser or Thr residue at position 7. This suggests a hierarchy in post-translational modifications. It is notable that in κA-SIVA and all homologous peptides (κA-conotoxin MIVA and the *Conus consors* peptide), position 7 is a serine or threonine residue that is always glycosylated, whereas Thr residues at positions 10 and 11 are not glycosylated (17, 27). κA-conotoxin MIVA is the first member of this family characterized that appears to have two glycosylated Thr residues.

The long N-terminal extension before the first Cys residue and the extensive suite of post-translational modifications present in κA-conotoxins further differentiate this branch of the A-superfamily from the α-conotoxin family. Thus, the A-conotoxin superfamily comprises at least two functionally differentiated (as well biochemically divergent) groups of peptides, the α-conotoxin family and the κA-conotoxin family. The former are targeted to nicotinic acetylcholine receptors, which belong to the ligand-gated channel superfamily, whereas the latter affect members of the voltage-gated ion channel superfamily.

Polymorphism of Conotoxin Genes—The results of this study provide insight into conotoxin polymorphism at several levels. The majority of the sequences encoding the 16 different conotoxins are likely either different A-superfamily genes, or homologous genes from different *Conus* species. However, there are quite a few cases in which different sequences may be allelic variants, including α-conotoxins GI/GIB, κA-conotoxins SIVA/SIVB, κA-conotoxins SmIVA/SmIVB, α-conotoxins SI/SII, and Bn1.1/1.2. For several of these possible allelic pairs

TABLE V

Some differences between natural toxin sequences and predicted cloned sequences

Underlined amino acids are *O*-glycosylated; γ, gamma-carboxyglutamate; O, hydroxyproline. α-Conotoxins SI and SII and κA-conotoxin SIVA were previously isolated by Zafaralla *et al.* (18), Ramilo *et al.* (32), and Craig *et al.* (17), respectively.

SI	
Precursor structure	---DRKEI**CC**NPA**C**GPKYS**C**GR
Predicted mature toxin	EI**CC**NPA**C**GPKYS**C**-NH_2
Isolated toxin	I**CC**NPA**C**GPKYS**C**-NH_2
SII	
Precursor structure	---NGRG**CC**NPA**C**GPNYG**C**GTS**C**SRTL
Predicted mature toxin	G**CCC**NPA**C**GPNYG**C**GTS**C**SRTL
Isolated toxin	G**CCC**NPA**C**GPNYG**C**GTS**C**S
SIVA	
Precursor structure	---HERQKSLVPSVITT**CC**GYDPGTM**C**PP**C**RC**T**NS**C**G
Predicted mature toxin	QKSLVPSVITT**CC**GYDPGTM**C**PP**C**RC**T**NS**C**-NH_2
Isolated toxin	ZKSLVPSVITT**CC**GYDOGTM**C**OO**C**RC**T**NS**C**-NH_2
MIVA	
Precursor structure	---HERAPELVVTATTN**CC**GYNPMTI**C**PP**C**MC**T**YS**C**PPKRKPGRRND
Predicted mature toxin	APELVVTATTN**CC**GYNPMTI**C**PP**C**MC**T**YS**C**PPKRKPGRRND
Isolated toxin	AOγLVVTATTN**CC**GYNOMTI**C**OO**C**MC**T**YS**C**OOKRKO-NH_2

(such as κA-conotoxins SmIVA/SmIVB), there are a few amino acid substitutions when the two mature sequences are compared; these mostly arise from single non-synonymous nucleotide substitutions. For the α-conotoxin GI/GIB pair, there are two additional C-terminal amino acids in α-GIB compared with α-GI, but the two sequences are otherwise identical; this could have occurred either by the insertion of six nucleotides to the α-GI gene, or the deletion of six nucleotides from the α-GIB gene (a "block" insertion or deletion).

The most drastic difference in mature toxin sequences is between κA-conotoxins SIVA and SIVB; in this case, an additional 11 amino acids are present in SIVB compared with SIVA. As is demonstrated in Table II above, a block of 25 nucleotides in one gene has been substituted by a non-homologous block in the other gene (a "block substitution"). In regions 3′ or 5′ to the substituted blocks, the nucleotide sequences are identical in SIVA and SIVB. It is noteworthy, however, that κA-conotoxin SIVB was only found in *C. striatus* specimens collected from the Hawaiian Islands; these have a characteristically different shell morphology from Philippine specimens assigned to this species. It is interesting that, although a difference in Hawaiian and Philippine specimens was observed with respect to the κA-conotoxins expressed, no such differences were found for the α-conotoxins: identical α-conotoxin SI and SII cDNA clones were sequenced from both populations of *C. striatus*, and no sequence differences were observed between the two populations for these α-conotoxins.

Branches of the A-superfamily—The most unexpected result from this work was the discovery that peptides belonging to the A-superfamily do not have a single arrangement of cysteine residues in the mature conotoxin region, a general feature of other conotoxin superfamilies (1, 28). Thus, the O-superfamily and the M-superfamily have the Cys patterns *XCXXCXCCX-CXXCX* and *XCCXCXXCXCCXX*, respectively. α-Conotoxin GI and all other α-conotoxins have the cysteine residues arranged in the primary sequence as follows: *XCCXXXCXXXCX* (the Class I/II Cys pattern, referring to the formal nomenclature for conotoxins). In all known natural peptides with this arrangement, the disulfide connectivity is Cys^1-Cys^3, Cys^2-Cys^4, which we refer to as the A-1 scaffold (1). Our results establish that a second group of peptides, with a different Cys pattern, *XCCXX-CXXCXCXXCXX* (a Class IV Cys pattern), also belongs to the A-superfamily; this is typified by κA-conotoxin SIVA. Given these differences in Cys patterns, α-conotoxins and the κA-conotoxins would be expected to have quite different scaffolds, and therefore divergent polypeptide backbone structures. Because the latter peptides are excitotoxins and not nicotinic antagonists, the two groups are therefore both functionally and structurally divergent.

However, at a genetic level, there is also a clear-cut division between two branches of the α-conotoxin family, the α4/7 subfamily (consensus sequence $XCC(X_4)C(X_7)CXX$) and the α3/5 subfamily (consensus sequence $XCC(X_3)C(X_5)CXX$). It is noteworthy that, although *C. magus* and *C. striatus* belong to the same group of fish-hunting cone snails, the α-conotoxin MII precursor is not closely related at the sequence level of the prepro region to α-conotoxins SI and SII from *C. striatus* (nor to any of the κA-conotoxins); rather, in the first N-terminal 39 amino acids, α-conotoxin MII exhibits much less sequence divergence from two α-conotoxins predicted from *C. bandanus* and *C. caracteristicus* cDNA clones (see Bn1.1 and Ca 1.2 in Tables IV and VI). If the first 18 aa of the propeptide region of these precursor sequences are compared, at least three coherent groups can be defined: the α3/5 conotoxin subfamily, the α4/7 conotoxin subfamily, and the κA-conotoxin family. All peptides that belong within the same group do not differ in sequence in the interval analyzed in Table VI by more than three amino acids; in contrast, if peptides across different groups are compared, they diverge by no *less* than 5 out of 18 amino acids in the region indicated in the table. Three peptides (Bn1.3, S1.1, and Ca1.1) do not belong to any of the groups above and are divergent from each other and may define additional subfamilies within the α-conotoxins.

Because the α4/7 subfamily is widely distributed in fish-hunting, worm-hunting, and mollusc-hunting species, it has been suggested that this may be the ancestral Cys pattern of the α-conotoxin family (1). The α3/5 subfamily is much more narrowly distributed; it is only found in two clades of fish-hunting *Conus* and is regarded as a specialization that may have arisen later in the evolutionary history of the genus. If the α3/5 subfamily evolved from an α4/7 ancestor, the arrangement of codons shown in Table IV for α-conotoxin SII suggests how the conversion from 4 aa in the first loop (of α4/7 conotoxins) to 3 aa in the first loop (of α3/5 conotoxins) may have occurred. If an α-SII-like peptide was generated from an α4/7 ancestor by mutation to a Cys codon immediately after the first two Cys codons in an α4/7-like peptide, this would decrease the loop size from four to three (presumably this would be accompanied by the addition of another cysteine residue, such as is found in α-conotoxin SII). Subsequent loss of the first cysteine codon

TABLE VI
Non-identity in propeptide region (aa 22–39)

The numbers indicate the number of non-identical amino acids for a pair of conotoxins. The horizontal and vertical lines between numbers delineate non-identities within a group and between groups of conotoxins.

Conotoxins	α-GI	α-SI	α-MII	bn1	cr2	κA-MIVA	κA-SIVA	κA-SmIVA
α-GI	–	3	7	7	8	9	9	9
α-SI/α-SII		–	6	6	7	7	7	7
α-MII			–	2	3	7	6	5
Bn1.1				–	2	8	7	7
Ca1.2					–	9	8	8
κA-MIVA						–	2	2
κA-SIVA							–	2
κA-SmIVA								–

Propeptide region		
α3/5	α-GI	---FPSERASDGRDDTAKDEG---
	α-SI	---FPSDRASDGRDDEAKDER---
	α-SII	---FPSDRASDGRDDEAKDER---
α4/7	α-MII	---FPSDRASDGRNAAANDKA---
	Bn1.1	---FASDRASDGRNAAAKDKA---
	Ca1.2	---FTSDRASEGRNAAAKDKA---
κA	κA-MIVA	---IPSDRASDGRNAVVH-ER---
	κA-SIVA	---IPSDRASDGRNAAVH-ER---
	κA-SmIVA	---IPSDRASDGRNAAVN-ER---

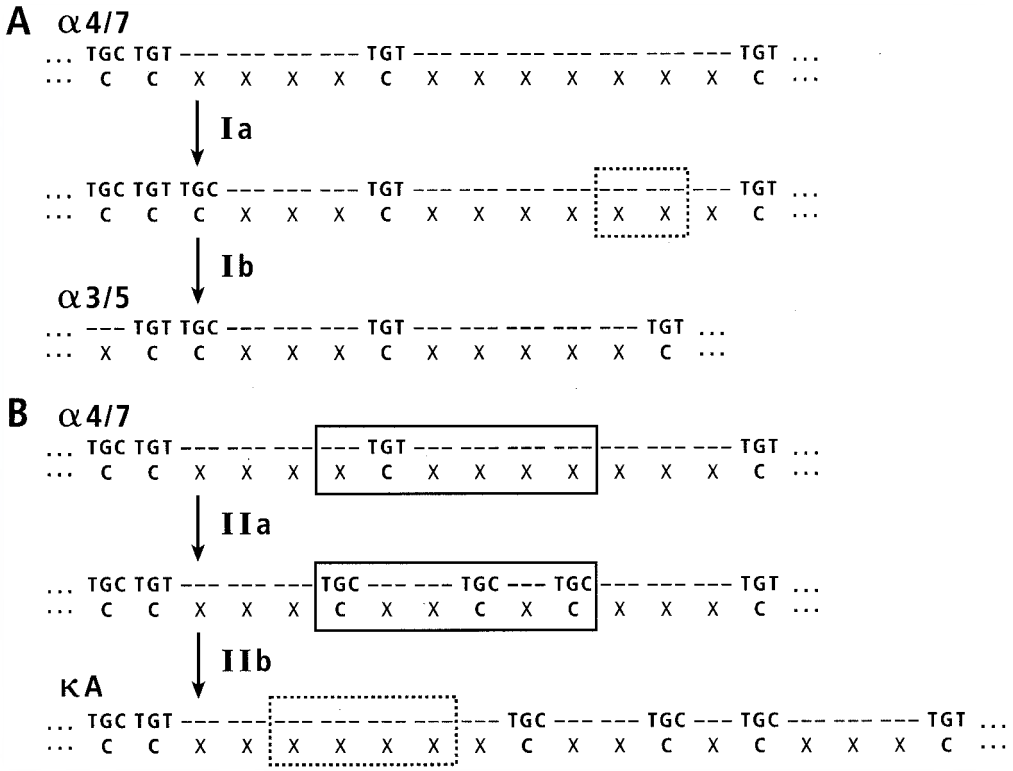

FIG. 3. **Proposed mechanism for the evolution of α3/5- and κA-conotoxins from a presumptive ancestral α4/7 conotoxin.** *A*, origin of α3/5 conotoxins. In *Pathway I*, the first step (*Ia*) involves mutation in the codon of the first amino acid in loop one of α4/7 to Cys (using a TGC codon). This is followed by two separate events (*Ib*), mutation in the first Cys residue to a non-Cys residue and the deletion of six nucleotides from loop 2 (*dotted box*). *B*, origin of κA-conotoxins. *Pathway II* consists of two steps: *IIa*, replacement of a block of nucleotides in the α4/7 ancestor (which includes the TGT codon for the third Cys residues) with another block of nucleotides containing the TGC codons for three Cys residues (*solid boxes*) to give the Cys pattern found in the κA-conotoxins, and *IIb*, insertion of a 12-nucleotide block (*dotted boxes*) in the first intercysteine loop.

(TGC in α-conotoxin SII and all α4/7 subfamily members) would then result in the codon pattern actually observed in α-conotoxins GI and SI (thus, the proposed transitions would be: TGC TGT NNN (the α4/7 pattern) → TGC TGT TGC (the SII pattern) → NNN TGT TCG (the α3/5 pattern, where NNN is a codon that does not specify Cys). The result would be a decrease in the size of the first loop of non-Cys aa from four to three, concomitant with an apparent reversal in the order of the Cys codons. This scheme is shown in Fig. 3. Thus, the codon reversal can be rationalized by the first Cys residue in α3/5 conotoxins being derived from the second Cys residue of the α4/7 conotoxins, both of which use the TGT codon.

In contrast, the κA-conotoxins have the conserved TGC TGT codons for the first two Cys residues as is generally found in the α-conotoxin family. However, Cys^3, Cys^4, and Cys^5 all are encoded by TGC codons in all κA-conotoxins analyzed; no TGC codons are used for Cys residues except for the first two in any α-conotoxins. One explanation consistent with this codon usage is

a "block substitution" mechanism, similar to that described above to explain the divergence between κA-conotoxins SIVA and SIVB. An ancestral κA-conotoxin could have been generated if a region containing Cys 3 of an α4/7 conotoxin ancestor were replaced by a block containing three Cys codons, all TGC (see Fig. 3 for a specific example of how this may have occurred).

In this study, we have defined the A-superfamily and delineated several major branches, one of which (the κA-conotoxins) is extremely divergent structurally and functionally. Moreover, the analysis of polymorphisms has uncovered genetic events (such as block substitution) that suggest how superfamily branches may have originated. Molecular mechanisms that underlie events such as block substitutions that drive toxin divergence are not known in detail at this time. It is known that A-superfamily genes have a large intron in the middle of the pro region[3]; it was previously suggested that the large introns found in conotoxin superfamily genes may play a role in the hypermutation observed in these peptides as speciation occurs (29). However, a detailed comparison of the introns from A-superfamily genes from different *Conus* species has not yet been made, and therefore this hypothesis has not been rigorously tested. Alternatively, block substitution events may arise from alternative splicing or retroviral-like template switching. However, the analysis presented here provides insight into potential transitional steps that occurred as structural and functional novelty evolved within a single conotoxin gene superfamily.

[3] R. Schoenfeld, B. Olivera, and D. Hillyard, unpublished results.

REFERENCES

1. Olivera, B. M. (2002) *Annu. Rev. Ecol. Syst.* **33,** 25–42
2. Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E. E., Woodward, S. R., Hillyard, D. R., and Cruz, L. J. (1990) *Science* **249,** 257–263
3. Kohn, A. J., and Perron, F. E. (1994) *Life History and Biogeography Patterns in Conus*, Clarendon Press, Oxford
4. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) *EMBO J.* **1,** 1015–1020
5. Olivera, B. M. (1997) *Mol. Biol. Cell* **8,** 2101–2109
6. Walker, C., Steel, D., Jacobsen, R. B., Lirazan, M. B., Cruz, L. J., Hooper, D., Shetty, R., DelaCruz, R. C., Nielsen, J. S., Zhou, L., Bandyopadhyay, P., Craig, A., and Olivera, B. M. (1999) *J. Biol. Chem.* **274,** 30664–30671
7. Lirazan, M. B., Hooper, D., Corpuz, G. P., Ramilo, C. A., Bandyopadhyay, P., Cruz, L. J., and Olivera, B. M. (2000) *Biochemistry* **39,** 1583–1588
8. Miles, L. A., Dy, C. Y., Nielsen, J., Barnham, K. J., Hinds, M. G., Olivera, B. M., Bulaj, G., and Norton, R. S. (2002) *J. Biol. Chem.* **277,** 43033–43040
9. Jimenez, E. C., Shetty, R. P., Lirazan, M., Rivier, J., Walker, C., Abogadie, F. C., Yoshikami, D., Cruz, L. J., and Olivera, B. M. (2003) *J. Neurochem.* **85,** 610–621
10. Gray, W. R., Luque, A., Olivera, B. M., Barrett, J., and Cruz, L. J. (1981) *J. Biol. Chem.* **256,** 4734–4740
11. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Press, Cold Spring Harbor, NY
12. Hultman, T., Stahl, S., Homes, E., and Uhlen, M. (1989) *Nucleic Acids Res.* **17,** 4937–4946
13. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
14. Cartier, G. E., Yoshikami, D., Gray, W. R., Luo, S., Olivera, B. M., and McIntosh, J. M. (1996) *J. Biol. Chem.* **271,** 7522–7528
15. Yoshikami, D., Bagabaldo, Z., and Olivera, B. M. (1989) *Ann. N. Y. Acad. Sci.* **560,** 230–248
16. Shon, K., Olivera, B. M., Watkins, M., Jacobsen, R. B., Gray, W. R., Floresca, C. Z., Cruz, L. J., Hillyard, D. R., Bring, A., Terlau, H., and Yoshikami, D. (1998) *J. Neurosci.* **18,** 4473–4481
17. Craig, A. G., Zafaralla, G., Cruz, L. J., Santos, A. D., Hillyard, D. R., Dykert, J., Rivier, J. E., Gray, W. R., Imperial, J., DelaCruz, R. G., Sporning, A., Terlau, H., West, P. J., Yoshikami, D., and Olivera, B. M. (1998) *Biochemistry* **37,** 16019–16025
18. Zafaralla, G. C., Ramilo, C., Gray, W. R., Karlstrom, R., Olivera, B. M., and Cruz, L. J. (1988) *Biochemistry* **27,** 7102–7105
19. McIntosh, J. M., Santos, A. D., and Olivera, B. M. (1999) *Annu. Rev. Biochem.* **68,** 59–88
20. Shon, K., Koerber, S. C., Rivier, J. E., Olivera, B. M., and McIntosh, J. M. (1997) *Biochemistry* **36,** 15693–15700
21. Zoli, M., Moretti, M., Zanardi, A., McIntosh, J. M., Clementi, F., and Gotti, C. (2002) *J. Neurosci.* **22,** 8785–8789
22. Champtiaux, N., Han, Z.-Y., Bessis, A., Rossi, F. M., Zoli, M., Marubio, L., McIntosh, J. M., and Changeux, J.-P. (2002) *J. Neurosci.* **22,** 1208–1217
23. Espiritu, D. J. D., Watkins, M., Dia-Monje, V., Cartier, G. E., Cruz, L. J., and Olivera, B. M. (2001) *Toxicon* **39,** 1899–1916
24. Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., and Fainzilber, M. (2001) *Mol. Biol. Evol.* **18,** 120–131
25. Eipper, B. A., Perkins, S. N., Husten, E. J., Johnson, R. C., Keutmann, H. T., and Mains, R. E. (1991) *J. Biol. Chem.* **266,** 7827–7833
26. Ping, D., Mounier, C. E., and May, S. W. (1995) *J. Biol. Chem.* **270,** 29250–29255
27. Favreau, P., Krimm, I., Le Gall, F., Bobenrieth, M.-J., Lamthanh, H., Bouet, F., Servent, D., Molgo, J., Ménez, A., Letourneux, Y., and Lancelin, J.-M. (1999) *Biochemistry* **38,** 6317–6326
28. Terlau, H., and Olivera, B. M. (2004) *Physiol. Rev.* **84,** 41–68
29. Olivera, B. M., Walker, C., Cartier, G. E., Hooper, D., Santos, A. D., Schoenfeld, R., Shetty, R., Watkins, M., Bandyopadhyay, P., and Hillyard, D. R. (1999) *Ann. N. Y. Acad. Sci.* **870,** 223–237
30. Santos, A. D. (1993) *Precursor Structures of Some α-Conotoxins*, Ph.D. thesis, University of the Philippines
31. Bai-Song, L., Fang, Y., Dong, Z., Pei-Tang, H., and Cui-Fen, H. (1999) *Peptides* **20,** 1139–1144
32. Ramilo, C. A., Zafaralla, G. C., Nadasdi, L., Hammerland, L. G., Yoshikami, D., Gray, W. R., Kristipati, R., Ramachandran, J., Miljanich, G., Olivera, B. M., and Cruz, L. J. (1992) *Biochemistry* **31,** 9919–9926
33. Olivera, B. M., Cruz, L. J., Hillyard, D. R., McIntosh, J. M. and Santos, A. D. (December 31, 1996) U.S. Patent 5,589,340