Novel αD-Conopeptides and Their Precursors Identified by cDNA Cloning Define the D-Conotoxin Superfamily[†,‡]

Marion L. Loughnan,[§] Annette Nicke,[||] Nicole Lawrence,[§] and Richard J. Lewis*[,§]

Institute for Molecular Biosciences, The University of Queensland, Brisbane, Qld 4072, Australia, and Max Planck Institute for Brain Research, Deutschordenstrasse 46, 60528 Frankfurt/Main, Germany

Received January 8, 2009; Revised Manuscript Received March 10, 2009

ABSTRACT: αD-Conotoxins are peptide inhibitors of nicotinic acetylcholine receptors (nAChRs) first described from *Conus vexillum* (αD-VxXIIA−C and renamed here to αD-VxXXA, αD-VxXXB, and αD-VxXXC). In this study, we report cDNA sequences encoding D-superfamily conopeptides identified in the Clade XII Conidae *Conus vexillum*, *Conus capitaneus*, *Conus mustelinus*, and *Conus miles*, together with partial sequences of corresponding peptides from this family. The D-superfamily signal peptide sequences display greater heterogeneity than reported for other conotoxin superfamilies. Phylogenetic analysis of the relationships among αD-conotoxin precursors reveals two distinct groups containing either an EMM or AVV signal peptide sequence motif. Homodimer and heterodimer combinations of predicted mature toxin sequences likely account for the partial amino acid sequences and mass values observed for several of the native dimeric peptide components identified in *C. capitaneus*, *C. miles*, and *C. mustelinus* venom. The discovery of the precursors and several novel conotoxins from different species defines this large conotoxin family and expands our understanding of sequence diversification mechanisms in *Conus* species.

Conotoxins are small disulfide-rich peptides found in the venom of predatory marine snails from the genus *Conus*. These peptide toxins generally target voltage- and ligand-gated ion channels (*1*). Nicotinic acetylcholine receptors (nAChRs)[1] are ligand-gated ion channels formed from homopentameric or heteropentameric combinations of α and β subunits (α2−α10 and β2−β4) producing a diversity of receptor subtypes with distinct physiological functions and pharmacological properties (*2*). Conotoxins that act at nAChRs include the α-conotoxin, αA-conotoxin, ψ-conotoxin, αS-conotoxin, αC-conotoxin, and αD-conotoxin families (*3, 4*). The common α-conotoxins are well-characterized, competitive nAChR antagonists with diverse subtype selectivities (*2, 5*). The αA-, αS-, and αC-conotoxins are competitive antagonists of the muscle nAChR subtype (*4*−*7*). In contrast, ψ-conotoxins are noncompetitive antagonists of muscle nAChRs (*6*) and αD-conotoxins VxXIIA, VxXIIB, and VxXIIC (here renamed VxXXA, VxXXB, and VxXXC)[2] are noncompetitive antagonists of neuronal nAChRs with selectivity for α7- and β2-containing subtypes (*3*). The unusually large αD-conotoxins contain ten conserved cysteine residues and occur as dimers of paired 47−50 residue peptides possessing varying levels of posttranslational heterogeneity associated with proline and glutamic acid conversion to hydroxyproline and γ-carboxyglutamic acid (*3*).

Most conopeptides are initially translated as prepropeptides that are posttranslationally processed to generate mature toxins (*8*). The precursors usually have a canonical architecture comprising an N-terminal signal sequence followed by an intermediate processive region and a single C-terminal mature toxin region (*9*). For precursors of large conotoxins containing three or four disulfide bonds, these three regions are encoded by three discrete exons separated by large introns (*9, 10*). Precursors of conotoxins from a single family display a characteristic transition from high sequence conservation at the signal peptide region through to hypervariability of the amino acids in the inter-cysteine loops of the mature toxin region, corresponding to differences in rates of codon substitution in each of the exons (*10*−*12*). To preserve the integrity of the assorted disulfide-bonded scaffolds in conotoxin classes, cysteine residues in the mature toxin are protected from hypervariation common across many of the non-cysteine residues in the mature peptides, probably

[†] M.L.L. was the recipient of an APA. This work was supported by an NHMRC Program Grant.

[‡] Nucleotide sequence data are available in the DDBJ/EMBL/GenBank databases under the accession numbers FN178633−FN178636. Protein sequence data for VxXIIA, VxXIIB, and VxXIIC (here renamed VxXXA, VxXXB, and VxXXC) are available in the UniProt database under the accession numbers P0C1W5, P0C1W6, and P0C1W7, respectively.

* Address correspondence to this author. Telephone: +61 7 3346 2984. Fax: +61 7 3346 2101. E-mail: r.lewis@imb.uq.edu.au.

[§] The University of Queensland.

[||] Max Planck Institute for Brain Research.

[1] Abbreviations: nAChR, nicotinic acetylcholine receptor; ESI-MS, electrospray ionization mass spectrometry; LC/MS, liquid chromatography mass spectrometry; MS, mass spectrometry; MS/MS, tandem mass spectrometry; PTM, posttranslational modification; TIC, total ion current; UP, universal primer.

[2] The peptides characterized in Loughnan et al. (*3*) were named αD-VxXIIA, αD-VxXIIB, and αD-VxXIIC. Overlap during publication resulted in another cysteine framework being given the identifier XII. In this article, we propose that conopeptides with the cystine pattern C-CC-C-CC-C-C-C-C have the identifiers 20 or XX, and we rename the original αD-conotoxins from *C. vexillum* as VxXXA, VxXXB, and VxXXC. See Materials and Methods for further detail.

10.1021/bi9000326 CCC: $40.75
Published on Web 03/10/2009

through a mechanism involving position-specific cysteine codon preference (either TGT or TGC) (*9, 12*).

Conotoxins have been grouped structurally into superfamilies on the basis of similarities in precursor signal peptide sequences and mature toxin cysteine frameworks (*11, 13*−*15*). Conservation of the signal sequences of their precursors has been used to identify predicted peptides belonging to a particular conotoxin superfamily by cDNA cloning (*12*). Previously, we have described the discovery and characterization of three posttranslationally modified αD-conotoxins, now renamed VxXXA, VxXXB, and VxXXC, which occur as dimers and produce a slowly reversing block of α7 and α3β2 nAChRs as demonstrated by two electrode voltage clamp analysis of oocyte-expressed nAChR subtypes (*3*). The sequences of these αD-conotoxins from *Conus vexillum* were determined by a combination of classical Edman N-terminal sequencing and tandem mass spectrometry. In the present work, we determined the cDNA sequences of 19 αD-conopeptide precursors that define the D-superfamily of nAChR antagonist conotoxins. We show that their signal sequences are less conserved, while the mature toxin regions are less variable, than is typically found in the conotoxin families characterized previously. The relatively conservative nucleotide changes that have generated the αD-conotoxin precursor sequence variants are indicative of positive selection and provide information about hypervariation processes in conotoxin biosynthesis.

MATERIALS AND METHODS

Crude Venom Extraction. Specimens of *C. vexillum* (eight), *Conus mustelinus* (seven), *Conus miles* (nine), *Conus capitaneus* (ten), *Conus rattus* (eleven), and *Conus vitulinus*[3] (twelve) were collected from the Great Barrier Reef, Australia. Crude extracts were prepared from venom duct material using 30% acetonitrile/water acidified with 0.1% TFA. Soluble crude venom extract was lyophilized and stored at −20 °C.

αD-Conotoxin Isolation. Crude venom extract (10 mg) was fractionated by semipreparative RP-HPLC (10 μm C18; Vydac) eluted at 3 mL/min with a linear gradient of 0−90% solvent B over 80 min using a Waters 600 solvent delivery system (A, 0.1% TFA, aqueous; B, 90% acetonitrile/0.09% TFA, aqueous). Activity at the nAChR (*3*) was further fractionated by analytical C18 RP-HPLC using 5 μm 4.6 × 250 mm Jupiter and Vydac columns eluted at 1 mL/min or a 3.5 μm 300 SB-C18 2.1 × 150 mm Zorbax column eluted at 0.18 mL/min using 0.5% or 1% linear gradients from 0 to 50% B over 45 min at 65 or 23 °C. To confirm activity was associated only with the 11 kDa components, crude venom was also fractionated by size exclusion chromatography (Superdex peptide HR10/30; Pharmacia) eluted with 30% acetontrile/0.1% TFA at 0.5 mL/min. The activities of partly purified proteins from *C. capitaneus*, *C. mustelinus*, and *C. miles* were confirmed by retesting the fractionated material by TEVC recordings on oocyte-expressed nAChRs, as previously described (*3*).

LC/MS Analyses of Crude Venoms and Toxins. LC/MS analyses of crude venom extracts or partly purified mixtures of dimeric toxins were undertaken on a PESciex API QSTAR Pulsar instrument equipped with an electrospray ionization source and linked to an upstream Agilent 1100 series HPLC system. LC was performed with a Zorbax C3, 2.1 × 150 mm, 5 μm RP-HPLC column eluted with 0−60% B over 60 min (A, 0.1% formic acid; B, 90% acetonitrile/0.09% formic acid, aqueous) at 180 μL/min and 65 °C. TOF-MS scans were run in positive ion mode over a mass range of 500−2200 amu with an ion spray voltage of 5300 V. Processing of LC/MS data was performed using the software package Bioanalyst (PE-Sciex, Canada). Apex mass is defined as the mass of the isotope distribution at maximum intensity as identified by the mass reconstruction tool in the Bioanalyst software.

Electrophoresis of Toxins. SDS−PAGE (Laemmli method) was run with either cast 16% polyacrylamide or precast 10−20% Novex tricine minigels (Invitrogen) under denaturing nonreducing conditions. Gels were either stained with colloidal Coomassie Blue G-250 (Bio-Rad Laboratories Inc., Hercules, CA), or proteins were transferred to PVDF using SDS/glycine buffers and then stained with Coomassie Blue R-250 (Bio-Rad).

Reduction and Alkylation of Cystine Residues. The semipurified ∼11 kDa proteins (∼20 pmol) were reduced in the presence of 10 mM TCEP, 50 mM ammonium acetate, pH 4.5, and 10% acetonitrile (37 °C for 1 h) and subsequently alkylated in the presence of 20 mM maleimide (37 °C for 1 h). The alkylated peptides were repurified by RP-HPLC.

Peptide Sequencing by Edman Chemistry. Edman N-terminal sequence analysis was undertaken on partly purified material with disulfide bonds intact or with cysteine residues alkylated with maleimide after reduction with TCEP, as described above. Sequence analysis was undertaken using an Applied Biosystems procise protein sequencer (HT or 492cLC models).

cDNA Preparation. Venom duct tissue from specimens of *C. vexillum*, *C. capitaneus*, *C. miles*, *C. mustelinus*, *C. rattus*, and *C. vitulinus* was finely diced, placed in RNAlater reagent (Ambion Inc., Austin, TX), and stored at −80 °C. Total RNA was extracted from preserved venom duct tissue (5−10 mg) using the RNeasy Mini RNA purification kit (QIAGEN Inc., Valencia, CA) and reversely transcribed using the SMART RACE cDNA amplification kit (Clontech Laboratories, Inc., Mountain View, CA) and Superscript reverse transcriptase (Invitrogen Corp., Carlsbad, CA). The reaction mixtures were incubated in a thermal cycler at 42 °C for 1.5 h, diluted with 40 μL of TE buffer, heated at 72 °C for 7 min, and then stored at −20 °C. Amplification of cDNA by LD PCR was carried out with the SMART PCR cDNA synthesis kit (Clontech Laboratories, Inc., Mountain View, CA) using the 5′ PCR primer IIA and Advantage 2 polymerase mix. Thermal cycler conditions were as follows: 94 °C, 30 s; 68 °C, 4 min (18 cycles). The resulting products were used as templates for gene-specific amplifications.

cDNA Cloning. The amino acid sequence of VxXXA (*3*) was used to design degenerate primers for gene-specific primer PCR amplification (GSP-PCR) of cDNA corresponding to a portion of the mature peptide region of the VxXXA precursor. The forward gene-specific primer (GSP) P1 (5′-GAYGTNCARGAYTGYCARGT-3′) was based on the motif DVQDCQV from the VxXXA peptide. The reverse GSP P2 (5′-RTGRTANACRCARTARCARTG-3′) was based

[3] *Conus planorbis f. vitulinus* is an alternative description for *Conus vitulinus*.

on the motif HCYCVYH from VxXXA. PCR conditions were as follows: 94 °C, 30 s; 45 °C, 30 s; 72 °C, 3 min (40 cycles).

RACE (rapid amplification of cDNA ends (*16*)) analysis was undertaken for cDNA cloning of full-length αD-conotoxin precursors, including signal peptide and 3′-UTR sequence regions, from *C. vexillum*. The cDNA sequence corresponding to the mature toxin peptide sequence for VxXXA, determined in the previous experiment, was used to design primers for use in 3′- and 5′-RACE amplification of the toxin-encoding cDNA. The 5′-RACE reverse primer P3 (5′-GCACCTTCCCCATTTTGAG-3′) and the 3′-RACE forward primer P4 (5′-TCAAAATGGGGAAGGTGC-3′) were both based on the cDNA sequence encoding the motif SKWGRC from the VxXXA peptide. Similarly, primer pairs P3B and P4B and P3C and P4C were based on the motifs SPWGRC and STWGRC from VxXXB and VxXXC, respectively. RACE was performed using first strand cDNA transcribed with the SMART system (Clontech, Palo Alto, CA) and the respective primers, including the universal primer (UP). Thermal cycling conditions were as follows: 94 °C, 30 s; 60 °C, 30 s; 72 °C, 3 min (30 cycles).

RACE and GSP-PCR analyses were undertaken to clone cDNA of additional αD-conopeptide precursors from *C. vexillum*, *C. capitaneus*, *C. miles*, *C. mustelinus*, *C. rattus*, and *C. vitulinus*. The cDNA sequence corresponding to a portion of the signal peptide sequence identified from *C. vexillum*, (M)PKLEMM(L), was used to design the forward primer P5 (5′-GCCAAAACTGGAAATGATGC-3′) for use in 3′-RACE amplification of the corresponding toxin-encoding cDNA. GSP-PCR was undertaken using the primer P5 as above together with an additional gene-specific reverse primer P6 (5′-GCTGCTTTGTTAGAGCAAGGTC-3′) designed from a conserved section identified within the 3′-UTR of sequences determined in the previous experiment. 5′-RACE using the primer P6 was also undertaken. Finally, 3′-RACE was undertaken using a gene-specific forward primer P7 (5′-TTGAACAGRATAGCAGACAACCA-3′) designed from a conserved section identified within the 5′-UTR of sequences obtained in the previous experiments. GSP-PCR was also undertaken with P6 and P7 using conditions described above.

Amplified cDNA products were separated by agarose (1.2%) gel electrophoresis, purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA), and ligated into the pGEM-T Easy vector (Promega, Madison, WI). Ligation products were transformed into TOP10 *Escherichia coli* cells (Invitrogen, Carlsbad, CA). Plasmid DNA of 10−20 positive clones per experiment was sequenced using Big-Dye Terminator v3.1 reaction premix (Applied Biosystems, Foster City, CA) with the pUC/M13 forward primer or the SP6 (reverse) promoter primer. Sequencing reaction products were analyzed by the Australian Genome Research Facility (AGRF), and sequences were assembled and assessed using FourPeaks freeware software (*17*).

Sequence Analysis. The signal peptide cleavage sites in the prepropeptides were tentatively predicted using the program SignalP (*18, 19*). The propeptide cleavage sites were identified from the characterized mature toxin peptide sequences. Nucleotide and amino acid sequences of toxin precursors from other classes were collated for comparison with αD-conotoxins: The long eight-cysteine Na_v-targeting scorpion α-toxin precursor sequences were obtained from EMBL (accession codes U28659, AAG39642, AAG00580, AAG00581, AF15659, AF156168, U28660, AF164203, AF155363, AF151796, M27704, and AF155364) and (*20*). The δ-atracotoxin (AcTx) precursor sequences (identification codes XenFW299, XenFW90, TFWSa8, SDW22, XenFW331, and SDW42) were obtained from Wilson and King (unpublished data). The α-conotoxin and δ-conotoxin precursor sequences were obtained from the literature (*21*−*30*). Codon-based multiple sequence alignments were prepared manually with the assistance of Clustal W (*31*) in BioManager 2.0 provided by ANGIS (http://www.angis.org.au) or Geneious Align (*32*). Average pairwise percent similarity and percentage of identical sites were computed using Geneious 3.6.1 (*32*). Pairwise distance measures (numbers of sites, numbers of differences, and *p*-distance values) were computed from aligned nucleotide sequences using MEGA 4.0 software package programs, implementing codon models and applying pairwise deletion of alignment gaps and missing data (*33, 34*). Geneious (*32*) sequence displays were used to highlight peptide and gene sequence variations.

Nomenclature. Overlap in the naming of conotoxins has arisen with the increased rate of discovery of new sequences in recent years. To date, the toxin disulfide framework identifier XII or 12 has been applied to the αD-conotoxins (*3*) together with the atypically named eight-cystine peptide GlaTxX from the I-superfamily (*35*), the precursors of a number of six-cystine peptides from the M-3 or mini-M-superfamily (*36*), and unpublished sequences in protein databases (UniProt Accession numbers A6YR20−42, A7LI89−92). To avoid further confusion, we propose in this paper that αD-conotoxins from *C. vexillum* (αD-VxXIIA, αD-VxXIIB, and αD-VxXIIC) be renamed αD-VxXXA, αD-VxXXB, and αD-VxXXC, with framework identifier XX and precursor sequence identifier 20. The cDNA sequences and the corresponding predicted peptides have been named with the two letter abbreviation for the species name, a number designating the mature toxin disulfide framework (20 or XX for αD-conopeptides), an appended number distinguishing multiple cloned sequences discovered from any one species (*37*), and an appended letter indicating the clone character (N for nucleotide sequences, P for predicted peptide precursor sequences, and M for predicted mature peptide sequences) (*38*). Additional conopeptides with this framework would be designated with the number 20 (if the target is not yet known) or XX (when the target has been defined).

RESULTS

Identification of Native Proteins. (A) Activity Profiles. As previously indicated (*3*), crude venoms (at 50 μg of protein/mL) from the vermivorous cone snails *C. capitaneus, C. mustelinus*, and *C. miles* contained components with the ability to produce prolonged inhibition of agonist-evoked currents of α3β2 (100%, 97%, and 37% inhibition, respectively) and α7 (100% inhibition for each species) rat neuronal nAChRs expressed in oocytes. Venoms from *C. capitaneus* and *C. mustelinus* also inhibited the α4β2 (100% and 65% inhibition, respectively). An α3β2 nAChR assay of fractions generated by size-exclusion chromatography for each crude venom confirmed the presence of inhibitory components in

the 15−5 kDa fraction of these species just as for *C. vexillum* (*3*). Interestingly, the "small peptide" fractions where α-conotoxins typically elute were inactive for all species except *C. rattus* (data not shown). These activity profiles, and the absence of significant inhibition of the α4β4 and α3β4 subtypes, were similar to the activity of αD-conotoxins from *C. vexillum*.

(B) Protein Identification. Analysis of the venoms from *C. mustelinus*, *C. miles*, and *C. capitaneus* revealed ~11 kDa proteins that were more complex than those previously observed in *C. vexillum* venom. As a consequence, these proteins could not be purified to discrete peaks by RP-HPLC and showed varying levels of microheterogeneity in LC/MS analyses. The observed apex mass values of the ~11 kDa native toxins ranged from 10940−11130 Da for *C. capitaneus* components, 10280−10380 Da for *C. miles* components, and 10820−10940 Da for *C. mustelinus* components. The dominant ~11 kDa components were observed as 10959 and 10988 Da in *C. capitaneus* venom, 10858 and 10893 Da in *C. mustelinus* venom, and 10381 Da in *C. miles* venom. For *C. capitaneus* and *C. mustelinus*, the dominant components were each part of a series of proteins differing by 16 mass units. SDS−PAGE of the intact proteins in each species confirmed an apparent molecular size of 10−12 kDa (data not shown).

Partial protein sequences were determined by Edman N-terminal sequence analysis of intact proteins transferred to membrane from SDS−PAGE gels or from intact or reduced and alkylated proteins that were partially purified by RP-HPLC. Sequence analysis of selected 11 kDa peptides gave partial sequences DVRγCQVNTOGSSXGKCCM and DNEAγCQINTOGSXWGKCMLTRMCGPMCXARSG from *C. mustelinus*, DVQγCQVVTOGSKWGRCCLNRVCG and DAQγCQVVTOGSKWGRCCLNRVCG from *C. miles*, and DNEAγCQINTOGS and γVQγCQVDTOGS or EVQγCQVDTOGS from *C. capitaneus*. From these partial sequences, a consensus sequence xxxx(x)Cxxx-TxxSxWGxCC that corresponded to the N-termini of VxXXA (DVQDCQVSTOGSKWGRCC) and VxXXB (DDγSγCIINTRDSPWGRCC) was identified.

The partial Edman sequence data obtained for most intact peptide samples included a secondary sequence in addition to the primary sequence, indicating multiple coeluting or interacting components. These partial sequences resembled but could not be matched with absolute certainty to the individual components observed by MS. To obtain complete sequence data for these αD-conopeptides, precursor gene sequences were determined by PCR analysis of cDNA from *C. mustelinus*, *C. miles*, and *C. capitaneus*, together with *C. vexillum*, *C. rattus*, and *C. vitulinus*.

Cloning of αD-Conotoxin Precursor Sequences. (A) Determination of the Full-Length cDNA Sequence Encoding VxXXA from C. vexillum. The cloning strategy, including the derivation of primers, is shown in Figure 1. An initial cDNA amplification by PCR with degenerate toxin-specific forward and reverse primers P1 and P2, designed from the amino acid sequence of VxXXA, yielded the cDNA sequence encoding the toxin fragment. The derived sequence was then used to design primers P3 for 5′-RACE and P4 for 3′-RACE which, together with subsequent 3′-RACE using primer P5, designed from the cDNA sequence encoding a portion of the signal peptide, yielded the complete coding sequence of the precursor of VxXXA, including a 3′-UTR of ~900 bp (Figure 2).

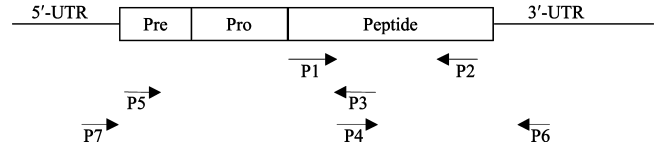

FIGURE 1: αD-Conotoxin cloning strategy. Diagram showing the strategy used to clone the sequences of αD-conotoxins from cDNA. The primers P1 to P4 refer to the cloning of VxXXA. A similar strategy was undertaken with primers P1b−P4b and P1c−P2c designed from VxXXB and VxXXC sequences. The primer sequences P1−P7 are given in Materials and Methods.

(B) Cloning of Additional αD-Conopeptide Precursors. C-Terminal sequence fragments of homologues of VxXXB and VxXXC precursors were also obtained from 3′-RACE with P4 and, together with the VxXXA precursor, revealed a conserved 3′-UTR sequence region. Primer P6 was designed from this sequence and used to obtain the complete coding sequence for VxXXB. Precursors for VxXXB and related variants were identified from *C. vexillum* cDNA by 5′-RACE using either the 3′-UTR-based primer P6 or the toxin-specific primer P3B. Precursors of VxXXA were also found in cDNA from *C. capitaneus* and *C. miles*, but not *C. mustelinus*, indicating that the same peptide may occur in the venom of at least three species. Interestingly, a precursor of VxXXB was also found in cDNA from *C. vitulinus*.

The cDNAs of precursors of several homologues of VxXXA, together with less closely related variants, were cloned using PCR undertaken with primers P5 and P6 or 3′-RACE undertaken with primers P5 and UP (universal primer). A second set of cDNA sequences obtained by cloning using 5′-RACE with the primers UP and P6 corresponded to precursors of homologues of VxXXB, together with additional novel sequences. Finally, a conserved region identified in the 5′-UTR of the cDNA sequences of precursors of homologues of both VxXXA and VxXXB was used to design primer P7, and PCR was undertaken with gene-specific primers P7 and P6. Several novel prepropeptides, some with toxin sequences intermediate between VxXXA and VxXXB, were identified from among the 180 clones isolated and sequenced from cDNA of the six species investigated.

Nineteen αD-conotoxin precursor sequences were identified at least twice ($n = 2-16$ clones) from the six species (Figure 3). Multiple closely related variants of these, each with one or two amino acid substitutions, were also identified. A search of the NCBI and UniProt databases revealed that these prepropeptide sequences and the corresponding mature toxin sequences, apart from VxXXA−VxXXC (formerly VxXIIA−C), had not been previously described. The cloned sequences are named as described in Materials and Methods. Some sequences differ by only one, two, or three nonadjacent residue changes in the mature peptide region. For example, Ms20.5M is a closely related [M20L,T26P] variant of Ms20.4M ($n = 2$), whereas Ms20.4M ($n = 3$) is a more distinct [V2N,R3E,-4A,V8I] variant of Ms20.3M ($n = 4$). Cp20.1M and Ml20.4M (or [E4D,V8S,I36V,K41R]Ml20.1M) are identical to Vx20.1M. Several unnamed variants ($n = 1$) of Cp20.1M have substitution at single positions 4, 19, 36, or 41. These include [D4G]Cp20.1M, [V36A]Cp20.1M,

```
   1                                gacca accagctgct tttcttgcct     25
  26    acctgcggtc tactacacgc atactttgaa caggatagca gacaaccacg     75

  76  ATG CCA AAA CTG GAA ATG ATG CTG CTG GTT CTG CTG ATT TTC CCC 120
   1  M   P   K   L   E   M   M   L   L   V   L   L   I   F   P    15

 121  CTG TCC TAT TTT ATT GCG GCA GGT GGA CAA GTG GTA CAA GTG GAC 165
  16  L   S   Y   F   I   A   A   G   G   Q   V   V   Q   V   D    30

 166  CGG CGT GGT GAT GGA CTG GCT GGT TAC CTC CAA CGT GGT GAC AGA 210
  31  R   R   G   D   G   L   A   G   Y   L   Q   R   G   D   R    45

 211  GAC GTT CAG GAC TGT CAA GTT AGT ACT CCA GGC TCA AAA TGG GGA 255
  46  D   V   Q   D   C   Q   V   S   T   P   G   S   K   W   G    60

 256  AGG TGC TGC CTG AAT CGG GTC TGC GGT CCC ATG TGC TGC CCT GCA 300
  61  R   C   C   L   N   R   V   C   G   P   M   C   C   P   A    75

 301  AGC CAT TGC TAC TGC GTC TAT CAT CGG GGA AGG GGC CAC GGG TGT 345
  76  S   H   C   Y   C   V   Y   H   R   G   R   G   H   G   C    90

 346  TCC TGC TGA tgagtatagct gctcagcgca ctctccgttc ttttttcctc    395
  91  S   C   *

 396    atgaccttgc tctaacaaag cagccaatca aagagcaacc gcaggtcctg    445
 446    aagcacgttt aaataaaacc aaccaaataa ttctatttag gactgctaat    495
 496    atatcttttt atctacatgt tcgaataaaa aaagatttgc attttacaaa    545
 546    tattttctat tttcactttg atctgactga gtcgggattt gagaatgctt    595
 596    atcgtttcgc ctcttagagt gcgaaggata caattaaagc aaaattggat    645
 646    cgcatttaaa agaaaacatt accgaaggga tatgctcaaa tggaaagata    695
 696    ctagagatag acaagacctc acctcatgca catgcatgta tagactaagt    745
 746    cctgtgacaa tattcataga gaagaaggac tttgtataaa tacacagaca    795
 796    aacattccgc acactttcca agaaaacgta acccaaagag cactttatat    845
 846    gccttttca gtttagctta taaagtagta cactccgaac agcactttta    895
 896    atggcctaaa atacagtgcc caaatgaaac cttttcaata ttcagcgata    945
 946    tagtagtgca ttttaagat ctttatatta ggagatcaga tagcattttt    995
 996    tttttcaaaa tatttttgt gaactttgtt tttttatatt tgttccgtgg   1045
1046    ttaatttggg aaatgctaag actaatttcg attacaaagt tatttgttcc   1095
1096    atcctacata aacaaccttg gaattttaga tgcagttttg ttcccccaat   1145
1146    gtggaagtgt tttattaaaa tgaattgtgc cgttcaagag aattaaagta   1195
1196    taaactgttt caatcaagag tatcgcattg ttttggacat gcagacggct   1245
1246    tgtgttagac tactgtagag aataaaaccg gcaataacat gccttgtact   1295
1296    g <poly A>
```

FIGURE 2: αD-VxXXA cDNA nucleotide sequence. The complete cDNA nucleotide sequence and predicted translation product for the precursor of αD-VxXXA. The signal sequence is highlighted (boxed region), the mature toxin region is underlined, and nucleotides in the UTR are in lowercase letters. The polyadenylation signal "aataaa" is boxed. The underlined sequence fragments in the 5′-UTR and 3′-UTR were used to design the forward primer P7 and the reverse primer P6, respectively.

[L19P]Cp20.1M, and [R41G]Cp20.1M, with predicted linear mature peptide mass values of 5076.09, 5106.09, 5118.07, and 5035.02 Da, respectively, compared with 5134.10 Da for VxXXA. Vx20.2M and vt20.1M are identical, although their precursors differ, and rt20.2M is a [V36I] variant of rt20.1M.

(C) αD-Conopeptide Precursor Structure. The identified sequences share the characteristic features of conotoxin precursors, including a moderately conserved signal peptide sequence region at the N-terminus, a "pro" region, a variable mature toxin region, and a closing 3′-UTR (see Figures 2 and 3). The predicted signal peptide sequences comprised 23−24 residues, the pro regions comprised 21 residues, and the mature toxin regions comprised 47−49 residues followed by stop codons. The cDNA sequences each had an unusually long 3′-UTR of 780−900 bp, ending with a polyadenylation signal "aataaa" and a poly(A) tail. Boundaries of regions (indicated by arrows in Figure 3) were inferred by prediction of signal peptide cleavage sites and presumption of propeptide proteolysis sites. Predicted sites of signal peptide cleavage included the motifs GG/QV, GG/QA, SG/QS, and GE/QV. The site of proteolytic cleavage of the mature peptide from the propeptide, deduced by comparison with known native αD-conotoxin peptide sequences, was highly conserved among the sequences. The pro region ended in the motifs LQRGDR or LQRGGR in most of the precursors, with cleavage occurring at the second basic arginine residue. All αD-conopeptide precursors had a common cysteine pattern, in addition to the moderately conserved signal peptide sequence.

An analysis of the precursor amino acid sequences was made to evaluate potential evolutionary relationships. This revealed two distinct αD-conotoxin groups with signal peptide motifs MXKLEMM (where X = P or L) or MPKLAVV (Figure 3). Precursors of VxXXA (Vx20.1P) and five identical or closely related peptides (Cp20.1P and Ml20.1P−Ml20.4P) had the signal peptide motif MPKLEMM, while the precursors of VxXXB (Vx20.2P and the identical peptide vt20.1P) had the signal peptide motif MPKLAVV. In addition, precursors of three sequences intermediate between VxXXA and VxXXB (Cp20.4P, Ms20.1P, and Ms20.2P) had the VxXXA-type precursor

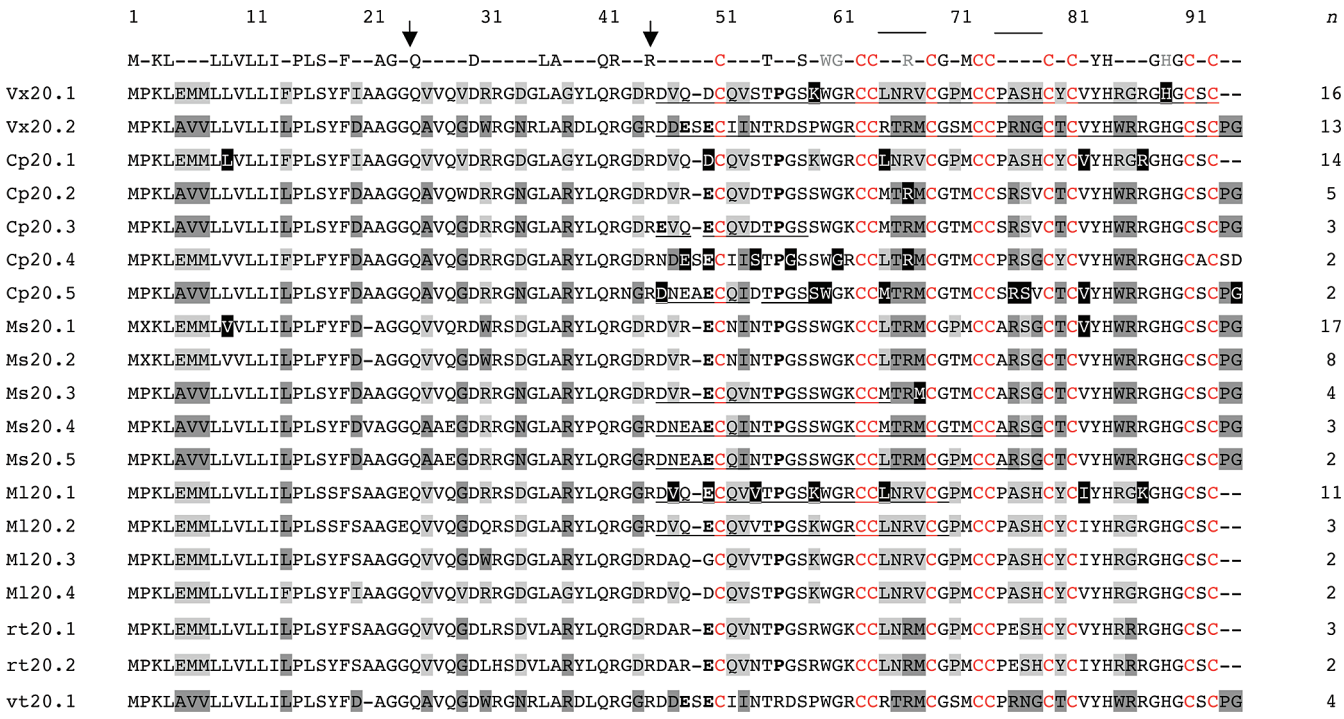

FIGURE 3: Prepropeptide amino acid sequences. Alignment of amino acid sequences of 19 prepropeptides from six species. Clones from cDNA prepared from venom ducts of *C. vexillum* (Vx), *C. capitaneus* (Cp), *C. miles* (Ml), *C. mustelinus* (Ms), *C. rattus* (rt), and *C. vitulinus* (vt) were obtained using PCR and RACE. Arrows indicate the predicted signal peptide cleavage sites (G/Q or E/Q) and mature peptide cleavage sites that separate preregions, proregions, and mature toxin regions of the precursor peptides. Regions of sequence that match observed N-terminal sequences of native peptides, excluding posttranslational modifications, are underlined. The Glu (E) and Pro (P) residues highlighted in bold have been modified in the mature peptides to Gla (γ) and Hyp (O). Single amino acids or motifs (particularly inter-cysteine loops 2 and 4, barred) that distinguish precursors of VxVIIA-like or VxXXB-like mature peptides are back-highlighted in light or dark gray, respectively. Selected isolated amino acids that are sites of high variation have been highlighted in white text on black background. The consensus sequence shows amino acids that are fully conserved (black) or partially conserved (gray). The identity of predicted mature peptides for each of the precursor sequences was as follows: Vx20.1M and Vx20.2M, identified from *C. vexillum*, correspond to VxXXA and VxXXB, respectively. Cp20.1M, identified from *C. capitaneus*, is identical to Vx20.1M. Cp20.2M, Cp20.4M, and Cp20.5M are novel sequences. Cp20.3M encodes a [D1E,R3Q] variant of Cp20.2M. Ml20.1M, identified from *C. miles*, is similar to Vx20.1M. Ml20.2M, Ml20.3M, and Ml20.4M encode variants equivalent to [K41R]Ml20.1M, [V2A,E4G,K41R]Ml20.1M, and [E4D,V8S,I36V,K41R]Ml20.1M, (identical to Vx20.1M), respectively. Ms20.1M, from *C. mustelinus*, is a novel sequence, resembling but not matching an observed peptide sequence. Ms20.2M is a [P25T] variant of Ms20.1M. The sequences Ms20.1 ($n = 11$) and Ms20.2 ($n = 6$) were obtained from 3′RACE using primer P5 (based on the Vx20.1 signal peptide motif MPKLEMM) or PCR using primers P5 and P6, but variants of Ms20.1 ($n = 6$) and Ms20.2 ($n = 2$) with the motif MLKLEMM were obtained from 5′RACE using primer P7 or PCR using primers P6 and P7, and thus the motif is shown as MXKLEMM. Ms20.3M and Ms20.4M are novel toxins that match the N-termini of two peptides identified from *C. mustelinus*. Ms20.5M ($n = 2$) is a [M20L,T26P] variant of Ms20.4M. The sequence vt20.1M, from *C. vitulinus*, is identical to Vx20.2M, although the precursors differ, and rt20.1M and rt20.2M, from *C. rattus*, are novel sequences with similarities to Vx20.M.

signal peptide motif MXKLEMM, as did rt20.1 and rt20.2. Another set of intermediate peptides had precursors (Cp20.2P, Cp20.3P, Cp20.5P, Ms20.3P, and Ms20.4P) with the VxXXB-type precursor signal peptide motif MPKLAVV. Finally, the precursor of a chimera (Vx20.3P), with a VxXXB-like N-terminal sequence and a VxXXA-like C-terminal sequence, also had the VxXXB-type MPKLAVV signal peptide motif.

(D) Assessment of Deduced Mature Peptide Toxin Sequences. The predicted mature peptide toxins were compared and where possible matched to the observed native peptides. Proposed heterodimer and homodimer combinations of predicted mature toxin sequences fitted the partial amino acid sequences and observed mass values (Figure 4) for some of the native peptide components from *C. capitaneus*, *C. miles*, and *C. mustelinus*. Six of the predicted sequences, Cp20.3M, Cp20.5M, Ml20.1M, Ml20.2M, Ms20.3M, and Ms20.4M, matched observed peptide N-terminal sequences for at least the first 18 residues (further Edman sequence data became unreliable) (see Figure 3). Mass values were calculated for the predicted sequences, taking into account presumed PTMs, and compared with the observed mass values of peptides (see Figure 4). The mixed N-terminal peptide sequence data, observed mass values of dimer components, and calculated mass values of predicted peptides together suggested that many of the αD-conotoxins from *C. capitaneus*, *C. miles*, and *C. mustelinus* were heterodimers that are interpreted further below. Peptides with sequences corresponding to Cp20.1M, Ms20.1M, rt20.1M, rt20.2M, and vt20.1 M remain to be identified in the venom of the corresponding species, despite being abundant at the cDNA level.

For *C. miles*, the observed mass values of ~10380.8 Da (intact) and 5206 Da (reduced), together with the partial Edman sequence data, suggested that one component in *C. miles* venom was a pseudohomodimer of [E4γ,P10O]Ml20.1M and [E4γ,P10O,P25O]Ml20.1M (calculated average mass 10384.1 Da for intact dimer; calculated monoisotopic mass 5206.16 Da for reduced [E4γ,P10O,P25O]Ml20.1M monomer). A second mass of 10656.6 Da remains unexplained.

For *C. capitaneus*, the precursors of numerous predicted αD-conotoxin variants were identified, and this complexity was reflected in the heterogeneity observed in LC/MS of

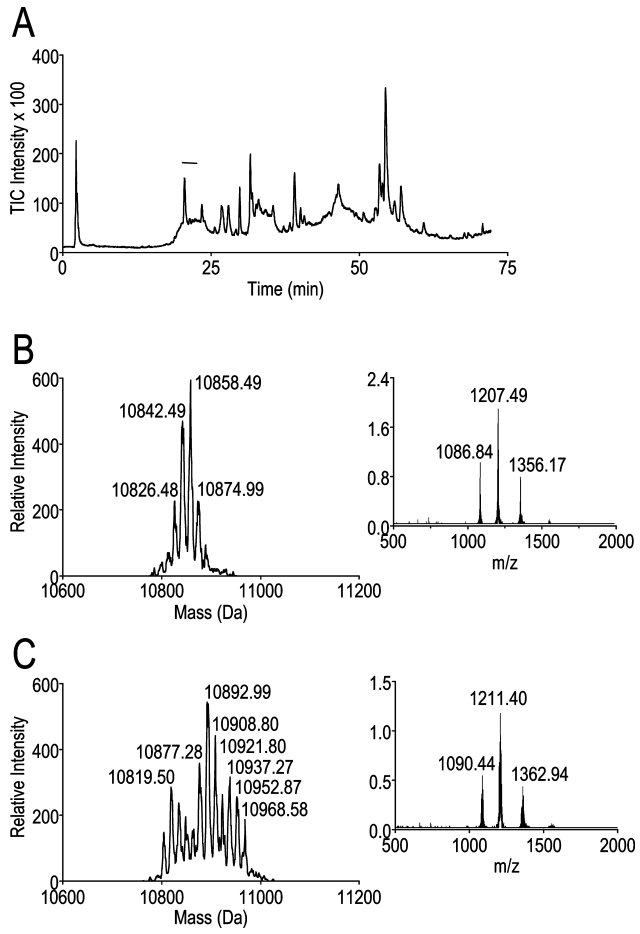

FIGURE 4: LC/MS of venom from *C. mustelinus*. (A) The TIC shows the elution time of peptides of interest, marked with a horizontal line, in LC/MS of a *C. mustelinus* crude venom extract. (B, C) Deconvoluted mass spectra are shown for some examples of 11 kDa peptides in the crude venom from *C. mustelinus*. Corresponding ion spectra are shown in the insets. The labels are maximum ion and mass values. The apex mass values for the dominant components in the crude venom were 10 858 and 10 893 Da, each part of a series of mass values differing by 16 mass units.

native peptides from this species. The N-terminal partial sequences observed for the peptides were consistent with dimeric combinations of posttranslationally modified variants of the predicted sequences Cp20.3M and Cp20.5M. Cp20.5M, with an N-terminal motif DNEAEC, resembles Ms20.4M. The existence of a further variant (identical to Ms20.4M) in *C. capitaneus*, either in a heterodimer of [E4γ,P10O]Cp20.3M and [E5γ,P11O]Ms20.4M or a homodimer of [E5γ,P11O]Ms20.4M, would explain some of the native peptide sequence combinations and mass values, but this variant remains to be identified in *C. capitaneus*.

In *C. mustelinus*, the predicted mature toxin sequences were consistent with two subsets of observed peptides with the N-terminal motifs DVRγC and DNEAγC. The 10842.49 Da mass and the first of the two peptide sequence fragments can be accounted for by a homodimer of [E4γ,P10O]Ms20.3M. The second partial peptide sequence resembled the Ms20.4M and Ms20.5M sequences. Heterodimer combinations of either [E5γ,P11O]Ms20.4M and [E4γ,P10O]Ms20.1M or [E5γ,P11O]Ms20.5M and [E4γ,P10O]Ms20.3M could account for the observed mass at 10877.28 Da.

(E) Toxin Precursor Sequence Comparisons. The cloning and peptide analyses establish that αD-conopeptides are more conservative in sequence than many other toxin classes. In order to better understand the distinctive features of codon usage in the αD-conopeptide precursors, this set of sequences was compared with four other sets of toxin sequences. The set of D-superfamily precursors displayed only a subtle gradation in protein sequence variability across the three regions from the signal peptide through the processing region to the mature toxin (Figure 5). This contrasts with the characteristic transition seen in precursors of many other classes of conopeptides, including the α- and δ-conotoxins, which show a highly conserved signal sequence and hypervariability of amino acids in the inter-cysteine loop regions of the toxin portion (see Figure 5). Differences in the extent of variation in each of the three regions of the conotoxin precursors were further illustrated in pairwise comparisons of two examples from each of five toxin families (see Supporting Information Table S1). The average pairwise percent similarity values for the nucleotide sequences of the mature peptide regions of the five sets of toxins were 86%, 65%, 65%, 67%, and 77% for the αD-CnTx, α-CnTx, δ-CnTx, δ-AcTx, and scorpion α-Tx sequence sets, respectively. The percentages of peptide region nucleotide sites that were identical in these sequence sets were 65%, 35%, 29%, 36%, and 48%, respectively. Thus at the nucleotide level, there was a higher proportion of conserved nucleotides in the mature peptide regions of αD-conotoxin precursors compared to the α- and δ-conotoxins (Figure 6) and the δ-AcTx and scorpion α-Tx precursor cDNA (data not shown). This relative conservation was also observed in a higher frequency of conserved amino acids among the noncysteine amino acids in the mature peptide region of the D-superfamily conopeptide precursors than typically observed for other conotoxin classes (see Figure 5). Moreover, the efficiency of variation was also higher, since a bias toward nonsynonymous substitutions resulted in more amino acid changes per nucleotide or codon change in the αD-conotoxin set than for other toxin classes (data not shown). Given this contrasting sequence conservation and codon divergence, the αD-conotoxin precursors were examined in further detail.

The predicted mature αD-conopeptides had the same cysteine framework (-C---CC---C---CC---C-C---C-C-) as found in αD-VxXXA, VxXXB, and VxXXC and shared another 12 conserved residues, namely, Thr10, Ser13, Trp15, Gly16, Arg22, Gly25, Met27, Tyr38, His39, Gly43, His44, and Gly45 (using VxXXB as the consensus sequence). Interestingly, VxXXC had only 6 of these 12 amino acids, and no VxXXC-related precursor sequences were obtained in this study, suggesting it may have arisen separately. Despite variation among the remaining 25−28 inter-cysteine residues, the predicted mature toxin sequences have similar theoretical p*I* values for the linear chains (7.7−8.6), and all contain 5−7 positively charged arginine or lysine residues, although their positions are not always conserved. Posttranslational modification sites (proline and glutamic acid) in VxXXA, VxXXB, or VxXXC were also not strictly conserved in the new αD-conotoxin sequences.

Preferred codon usage for many of the conserved amino acids in the αD-conotoxin precursor sequences was evident. Cysteine codon triplet usage was highly conserved for 8 of

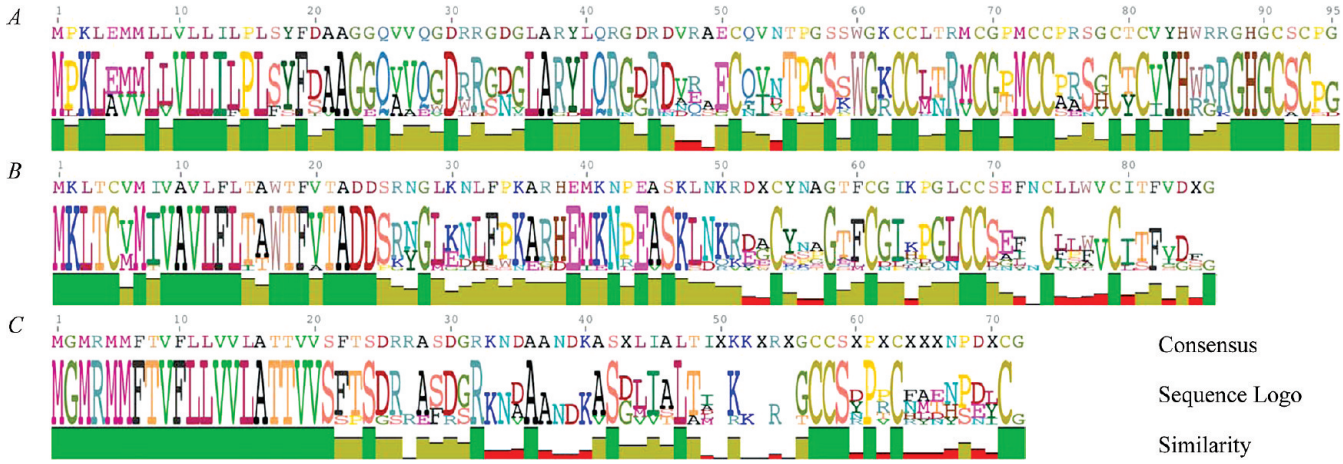

FIGURE 5: Comparison of precursor amino acid sequences from three conopeptide classes. Sequence logos, consensus sequences, and similarity indices (Geneious v3.2) derived from aligned precursor amino acid sequences from each of the (A) αD-, (B) δ-, and (C) α-conotoxin families. The analyses included precursors of 14 αD-conotoxins (Vx20.1, Vx20.2, Cp20.2, Cp20.3, Cp20.4, Cp20.5, Ml20.1, Ml20.2, Ml20.3, Ms20.1, Ms20.2, Ms20.3, Ms20.4, and Ms20.5), 12 δ-conotoxins (A6.5, Cn6.9, E6.4, Gm6.1, M6.3, M6.6, M6.7, M6.8, P6.5, S6.5, Sm6.6, and Tx6.16), and 14 α-conotoxins (ArIB, AuIA, AuIB, EpI, GIC, ImI, ImII, MII, PeIA, PIA, RgIA, SrIA, TxIA, and Vc1.1).

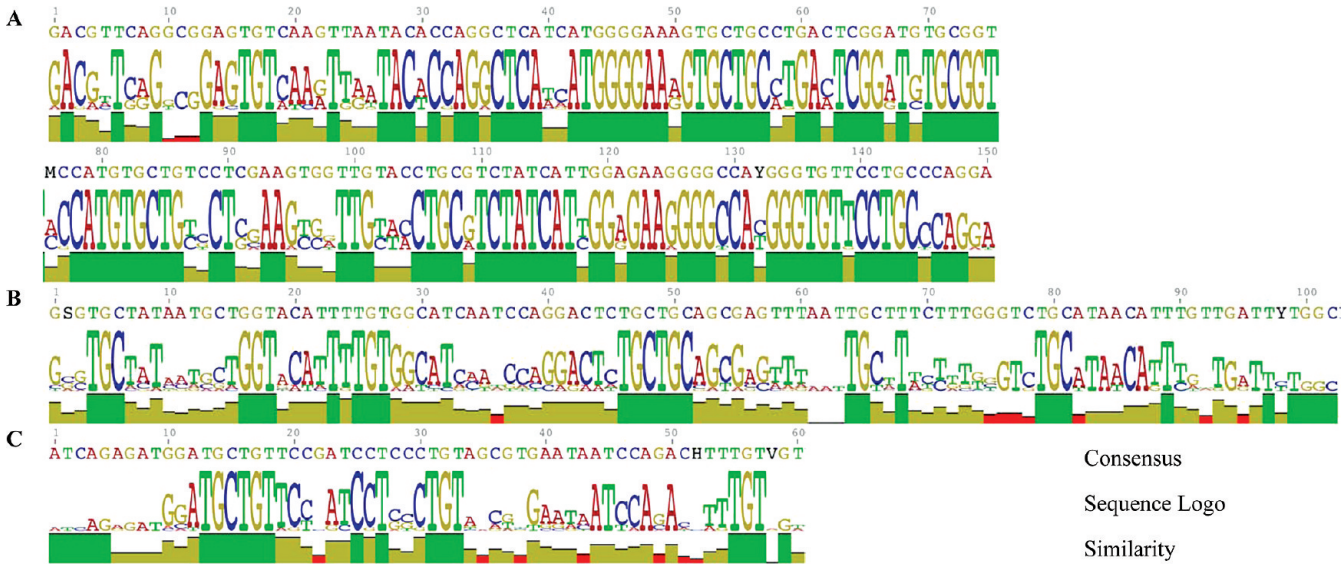

FIGURE 6: Comparison of precursor nucleotide sequences from three conopeptide classes. Sequence logos, consensus sequences, and similarity indices (Geneious v3.2) derived from aligned nucleotide sequences of mature peptide regions in precursors from (A) αD-, (B) δ-, and (C) α-conotoxin families. The sequences used in the analyses were as described for Figure 5.

the 10 cysteine positions, with TGC used for cysteine residues II, III, IV, V, VIII, and X, TGT used for cysteine residues I and IX, and TGT or TGC used for cysteine residues VI and VII. Interestingly, examination of the 19 predicted αD-conotoxin nucleotide and amino acid sequences revealed that codon preference was not restricted to cysteines but was also apparent for some other amino acids (Supporting Information Figure S1). Position-specific codon conservation was exhibited for 3 of the 4 fully conserved glycine residues (Gly16 GGA, Gly25 GGT, and Gly45 GGG, but not Gly43 GGC or GGT) and partly conserved glycine residues (Gly12 GGC and Gly33 GGT). There were no completely conserved proline or glutamate residues. One of the two possible codons (GAG and not GAA) was used for glutamate residues in the N-terminal positions 1, 3, or 5, whereas GAA was used for Glu31 in rt20.1M and rt20.2M. Where proline residues occurred, there was codon conservation at each position (Pro11 CCA, Pro14 CCA, Pro26 CCC, Pro30 CCT, Pro49 CCA, numbered relative to VxXXB). Additionally, the conserved amino acids Ser13, Arg22, Tyr38, and His39 also had conserved codons, whereas the conserved amino acids Thr10, Gly43, and His44 showed synonymous codon variation.

Codon usage for variable amino acids in the αD-conotoxin precursors was also distinctive. Examination of the cDNA sequences (see Supporting Information Figure S1) revealed patterns in nucleotide substitutions, with the frequency of changes in nucleotides varying with their position within the codon triplet (nt1 > nt2 > nt3). Across the set of αD-conotoxin precursor sequences, there were 20 nt1 changes, 18 nt2 changes, and 11 nt3 changes in a total of 150 nucleotide positions in the mature peptide regions, relative to a consensus sequence (see Supporting Information Figure S1). Thus there were relatively few synonymous changes. Analysis of a set of α-conotoxin precursor sequences revealed a similar trend, although with more frequent substitutions, with 12 nt1 changes, 9 nt2 changes, and 7 nt3 changes in up to 72 nucleotide positions in the mature peptide regions. With the high level of nt1 and nt2 changes, many amino acid changes among the αD-conotoxin precursors (with the exception of positions 7, 9, 14, 23, 31, 33, 35, and

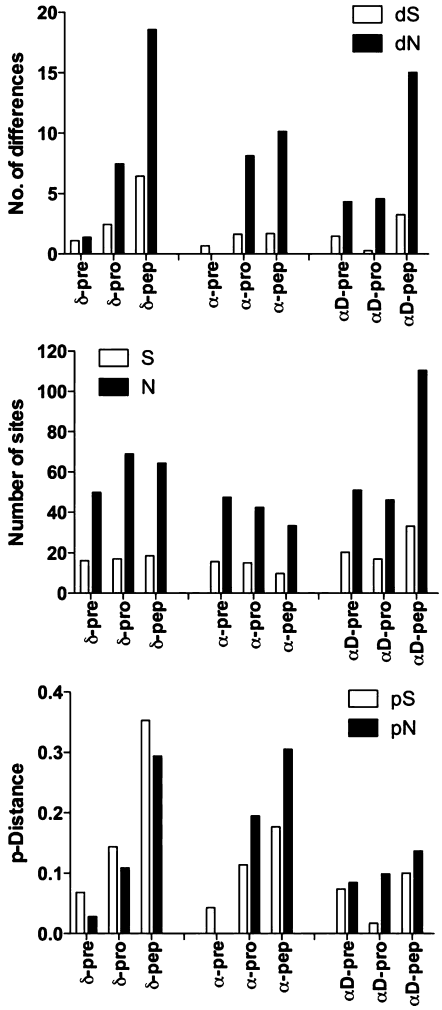

FIGURE 7: Comparison of substitution type across three precursor regions for each of three toxin classes. (A) The numbers of synonymous or nonsynonymous differences per sequence, calculated from pairwise comparisons, averaged over all sequence pairs. (B) The numbers of synonymous and nonsynonymous sites from averaging over all sequence pairs within each set of sequences. (C) The *p*-distance values, or numbers of synonymous changes per synonymous site, and nonsynonymous changes per nonsynonymous site, from averaging over all sequence pairs, determined for each of three regions of conotoxin precursors from three toxin classes (δ-, α-, and αD-conotoxins). The data were calculated using MEGA as described in Materials and Methods.

50, numbered relative to VxXXB) resulted from a single nucleotide change within the relevant codon (see Supporting Information Figure S1). In contrast to the relatively conservative changes in αD-conotoxin sequences, sets of sequences of other toxins, such as δ-conotoxins, δ-atracotoxins, and scorpion α-toxins, display more extensive nucleotide changes, both synonymous and nonsynonymous, and more amino acid changes. A comparison of codon substitution types revealed different rates of synonymous and nonsynonymous changes across the αD-, α-, and δ-conotoxin precursors for each of three precursor regions (Figure 7). The ω parameter (the ratio of the rate of nonsynonymous changes per nonsynonymous site to the rate of synonymous changes per synonymous site) was <1 for each of the δ-conotoxin, δ-atracotoxin, and scorpion α-toxin mature toxin region nucleotide sequence sets and >1 for the αD- and α-conotoxin sequence sets, suggesting only the latter two sets had the characteristics of positive selection. The high ω value for the αD-conotoxin sequences was attributable to particularly low synonymous changes. There was also a lower frequency of transversional nucleotide changes in the mature peptide regions of the αD-conotoxins than in the other four toxin classes (data not shown), so that the nonsynonymous changes in the former were predominantly transitional and hence conservative.

DISCUSSION

We previously identified the first αD-conotoxins VxXXA, VxXXB, and VxXXC (formerly VxXIIA, VxXIIB, and VxXIIC) with activity at nAChRs. The present study identifies D-superfamily sequences from six *Conus* species using a cDNA approach. The cDNA sequences of precursors of VxXXA, VxXXB, and additional novel predicted peptides from *C. vexillum*, *C. capitaneus*, *C. miles*, *C. mustelinus*, *C. rattus*, and *C. vitulinus* reveal the consensus nucleotide sequence of the untranslated and coding regions, the overall precursor structure, and the precursor signal sequences, thus defining the D-superfamily of conotoxins.

A single conotoxin gene superfamily has typically been characterized by representatives each having a highly conserved signal sequence region within the precursor sequence and a specific pattern of cysteine residues in the mature toxin region (*39*). The cysteine pattern and the consensus signal peptide sequence of the αD-conopeptide precursors are distinct from those of S-superfamily conopeptide precursors, which have 10 cysteines arranged in the class VIII cysteine pattern (*7, 40*). Among the αD-conotoxin precursors there are at least two signal peptide sequence variants, with either an MXKLEMM or an MPKLAVV motif. The signal peptide motif for precursors of VxXXC-like peptides remains to be identified. While the signal peptide sequences for α-conotoxin precursors from the A-superfamily are generally almost invariant (*28*), varied signal peptide sequences, which define multiple superfamily branches, have also been described for the O- (*12, 38, 41*), I- (*13, 42*), M- (*43*) and T- (*14, 44, 45*) superfamilies. We propose that αD-contoxins be placed in the D-1 branch of the D-superfamily to reflect their common cysteine pattern and consensus signal peptide motif.

For each of the αD-conotoxin precursors, the site of proteolytic cleavage of the mature peptide from the propeptide broadly fits the cleavage site motif (K/R)-$(X)_n$-(K/R)↓, where $n = 0$, 2, 4, or 6, a site recognized by subtilisin-type proprotein convertases (*46*). The αD-conotoxin precursors have the motif L-X-R-X_2-R↓ rather than the cleavage site motif L-X-(K/V/L/I)-R↓ seen in many conotoxin precursors (*47*). The resulting predicted toxin sequences match complete and partial sequences of peptides with nAChR-inhibitory activity isolated from *Conus* venom, after accounting for the PTMs that add another level of diversity to these conotoxins. The αD-conotoxins possess two common posttranslationally modified residues, hydroxyproline and γ-carboxyglutamic acid, and multiple variants of αD-conotoxins differ in the extent to which proline and glutamic acid residues are modified. A γ-carboxylation recognition site (γ-CRS), characterized by both hydrophobic and charged residues, has been postulated to occur within pro- or postpeptide regions of precursors of γ-carboxylated conotoxins (*35, 48, 49*), although such sites remain to be identified in the varied αD-conotoxin precursors. The novel predicted αD-conotoxins

Table 1: Conotoxin Nomenclature Showing Cysteine Arrangements[a]

cysteine arrangement	designation	superfamily	ref
CC-C-C[b]	1, 2 (I, II)	A, T	*67*
CC-C-C-CC	3 (III)	M	*68*
CC-C-C-C-C	4 (IV)	A	*69*
CC-CC	5 (V)	T	*70*
C-C-CC-C-C	6, 7 (VI, VII)	O	*71*
C-C-C-C-C-C-C-C-C-C	8 (VIII)	S	*40*
C-C-C-C-C-C	9 (IX)	P	*72*
C-C	10 (X)	C	*4*
C-C-CC-CC-C-C[c]	11 (XI)	I_1/I_2	*73*
C-C-C-C-CC-C-C	12 (XII)	I	*35*
C-C-C-CC-C-C-C	13 (XIII)	nd[g]	*74*
C-C-C-C[d]	14 (XIV)	J, L, nd	*75*
C-C-CC-C-C-C-C	15 (XV)	O[e]	*45*
C-C-CC	16 (XVI)	M[e]	*45*
C-CC-C-CC-C-C-C-C[f]	20 (XX)	D	this study

[a] The established nomenclature for naming conotoxin frameworks is based on the cysteine pattern alone (primary structure) regardless of disulfide connectivity or higher order structure. A separate scaffold classification takes into account disulfide connectivity and superfamily (*11*). [b] The conotoxin sequence MrIA has also been named mr10a (*76*), where framework 10 or X distinguished a different disulfide connectivity from that of framework I, but that naming has since been retracted (*4*) since the framework I naming for this sequence is consistent with the accepted nomenclature. [c] The cysteine pattern of the I_2 superfamily conotoxin κ-BtX, although also initially given the framework identifier X (*77*), is shared by other toxins that in the same year were named with the framework identifier XI, and the latter name has persisted. The framework identifier X has subsequently been used for the cysteine pattern C-C (*4*). [d] The peptides that share framework identifier XIV, or 14, are diverse, with representatives from different superfamilies indicated by variations in the signal peptide sequences of their precursors and differences in disulfide connectivity and size of inter-cysteine loops of the mature peptides. Nevertheless, they are placed together in category XIV on the basis of the cysteine pattern in the primary structure. [e] The peptides with framework identifiers XV and XVI have been categorized in subgroups of the divergent O- and M-superfamilies, respectively (*45*), although the signal peptide sequences may be considered sufficiently distinctive to place them in new superfamilies in the future. [f] The αD-conotoxins were initially given the toxin disulfide framework identifier XII or 12 (*3*), and we have proposed in this paper that they be renamed with framework identifier XX (leaving identifiers XVII, XVIII, and XIX available for renaming of other frameworks for which names have currently been duplicated). [g] nd: not defined.

from *C. capitaneus*, *C. miles*, and *C. mustelinus* assemble as both homodimers and heterodimers in contrast to the previously described VxXXA, VxXXB, and VxXXC, from *C. vexillum*, which occur as homodimers (*3*). This is reminiscent of the κ-neurotoxin nAChR antagonists from *Bungarus* snake venoms, which also occur as homodimers (*50, 51*) and heterodimers (*52*) in solution.

The predicted αD-conotoxin sequences have 47−50 residues, including 10 cysteine residues in a conserved pattern that defines the D-superfamily of conotoxins (Table 1). The importance of the fixed cystine framework for the structural integrity of mature toxins and the feature of position-specific cysteine codon conservation have been described for several families of conotoxins (*12, 38, 53*−*56*) and spider toxins (*9*). The αD-conotoxin precursors show codon preferences for cysteines and a number of other fully or partly conserved amino acids. Apart from these conserved elements, and the WG(R/K)CC and GHGC motifs, there is variability in the mature αD-conotoxin sequences, but the changes are relatively conservative and in many instances are attributable to a single nucleotide change in the relevant codon. The αD-conotoxins are thus unusual in displaying less extensive variation than many other toxin classes.

Venom peptides are suggested to be undergoing rapid evolution, resulting in divergent and hypervariable mature toxin sequences. Most evolutionary changes and variations in genes are neutral, with selection serving as a conservative force, while positive selection is considered to drive rapid adaptive evolution of some genes (*57*). Two types of natural selection have been distinguished: positive selection facilitates amino acid changes and promotes advantageous mutations while purifying selection restricts amino acid changes and eliminates deleterious mutations (*58*). Positive selection, characterized by higher rates of substitution of nonsynonymous nucleotides (causing amino acid change) than synonymous nucleotides (not causing amino acid change), has been identified in several multigene families that encode proteins with roles in offense, defense, and reproduction (*59*). Examples of such proteins in defense systems include those involved in acquired immunity (MHC, immunoglobulins) and innate immunity (transferrins, defensins, and other antimicrobial peptides), while examples of proteins in offensive systems in which positive selection has been recognized include toxins of cone snails, spiders, scorpions, and snakes (*57, 59*). Presumably, hypervariation in toxin sequences results from positive selection (i) to counter the development of prey resistance, (ii) to allow shifts in prey preference as prey abundance changes, and (iii) to facilitate niche diversification and speciation.

Genes at which positive selection is driving rapid adaptive evolution are unusual and show unique signatures that are detectable by phylogenetic analysis of aligned sequences using codon or amino acid evolutionary models (*57, 60*−*62*). The ratio of nonsynonymous to synonymous substitution rates (ω, Ka/Ks, or d_N/d_S) is indicative of selection type and intensity acting at specific positions. Sites with ω significantly lower than 1 are regarded as undergoing purifying selection and may have a functionally or structurally important role, whereas sites with $\omega > 1$ are regarded as experiencing positive Darwinian selection, suggesting adaptive evolution (*62*). Despite mature toxin sequences and other hypervariable peptides having been widely reputed to result from strong positive selection, here phylogenetic analyses of δ-conotoxins, δ-atracotoxins, and scorpion α-toxins have not shown the characteristic substitution pattern ($\omega > 1$) expected for positive selection. It seems that these toxins, unlike the αD-conotoxins, have sequence changes across nt1, nt2, and nt3 and do not show evidence of positive selection, suggesting that the evolution of the δ-conotoxins, δ-atracotoxins, and scorpion α-toxins is quite distinct from that of the αD-conotoxins. In contrast, the αD-conotoxins, while showing less overall variability than many other toxin classes, display the characteristic signature of positive selection, with a bias toward changes at positions nt1 and nt2 and contrasting conservation at nt3. The number of triplet codons encoding each amino acid is six for Arg, Ser, and Leu; four for Ala, Gly, Pro, Thr, and Val; three for Ile; two for Asn, Asp, Cys, Gln, Glu, His, Lys, Phe, and Tyr; and one for Met and Trp. The codon design symmetry and degeneracy of the third nucleotide position (nt3) for all amino acids except Arg, Ser, and Leu, for which there are six possible codons, mean that changes in the nt3 position are usually silent mutations, as in the example of Cys (TGT or TGC). It is often assumed that synonymous mutations are free from selection pressure and are not subjected to evolutionary forces (*53*), and codon

evolutionary models often assume that the purifying selection acting on protein-encoding DNA sequences is the result of selection that operates at the protein level only (*63*). However, it has been increasingly recognized that selection forces can also act at DNA or mRNA levels (*62, 64*). This has relevance in the search for mechanisms underlying the preservation of the cysteine frameworks and other highly conserved residues of conotoxins. The difference in frequency of nucleotide replacement at different positions in the codons of mature peptide regions suggests that the hypervariation mechanism in conotoxin genes preferentially targets positions nt1 and nt2, causing nonsynonymous changes, as in the example of the αD-conotoxins and to a lesser extent the α-conotoxins. This is consistent with some of the hypotheses for mechanisms of concurrent protection of conserved residues and hypervariation of intervening residues in conotoxins, including processes of mutagenesis through error-prone mismatch repair enzymes, mutator and antimutator polymerases, site-specific recombinases and integrases, and regional variations in stability of RNA secondary structure (*10, 54, 65*). The specific mechanisms associated with opposing hypervariability and conservation of conotoxin amino acids remain to be identified.

Gene duplication and focal hypermutation events followed by positive selection are proposed as a likely mechanism to generate hypervariability in conotoxins (*9, 20*). The αD-conotoxin precursors show evidence of recombination events, with differing combinations of mature toxin and signal sequence motifs seen for sequences intermediate between VxXXA and VxXXB precursors. Conotoxin gene architecture encodes pre, pro, and mature conotoxin regions by discrete exons separated by extensive introns (*10*). Given this arrangement, it is feasible that the exon encoding the mature toxin region can mutate at a different rate than the exons encoding the pre and pro regions of the conotoxin precursors (*10*). Identification of intron/exon boundaries and conserved intronic regions in genomic DNA sequences that encode αD-conotoxin precursors may shed further light on the hypermutation processes. The αD-conotoxin precursor sequences described here appear to be more closely related than those in some other conotoxin classes. However, it is unclear whether they are more recently evolved or conversely refined to a mature stage of evolution relative to other conotoxin classes. The biological significance of differences in substitution patterns within conotoxin sequence sets in response to evolutionary pressures is not known. The divergence in conotoxin superfamily sequences, both within and between *Conus* species, presumably allows target optimization and accommodates intra- and interspecies diversity of receptor subtypes in prey (*15*). *C. vexillum*, *C. mustelinus*, *C. miles*, *C. capitaneus*, and *C. rattus*, are Clade XII species, while *C. vitulinus* is a Clade IX species (*29*). Generally, conotoxins of most major superfamilies appear not to be clade-specific and instead are distributed widely among major *Conus* clades across different prey categories (*13*). However, examples of clade-specific strategies to target receptors include the use in three different clades of piscivorous cone snails of three different conopeptide families (κ, κM, and conkunitzins) for K^+ channel blockade (*66*). Neuronally selective αD-conotoxins appear to be restricted largely to Clade XII *Conus* species that lack α-, αA-, and αS-conotoxins, suggesting a key role for the αD-conotoxins in prey capture by this clade.

In conclusion, we describe the αD-conopeptide precursor protein sequences from four Clade XII species, *C. vexillum*, *C. mustelinus*, *C. miles*, and *C. capitaneus*, together with related gene sequences from *C. rattus* (Clade XII) and *C. vitulinus* (Clade IX). All members of the D-superfamily identified to date have the typical prepropeptide structure found in most other conotoxin precursors but display greater heterogeneity in the signal peptide regions than previously reported for other superfamilies. The αD-conotoxin sequence variants appear to have arisen from relatively conservative nucleotide substitutions through a mechanism that appears to target nt1 and nt2, generating nonsynonymous codon changes at hypervariable sites, while protecting nt3 to minimize silent synonymous codon changes, perhaps implicating nt3 as an anchor point for mutagenesis (*9, 54*). Thus the predicted αD-conopeptide sequences expand our understanding of the mechanisms contributing to toxin scaffold preservation and sequence divergence. Tertiary structural characterization of these peptides is required to start to unravel the structure−activity relationships among the diverse αD-conotoxins and their interactions with the neuronal nAChRs.

ACKNOWLEDGMENT

We thank Professor Glenn King and Dr. David Wilson (IMB) for kindly providing access to the set of sequences of δ-atracotoxins from spider venom used in comparisons with the αD-conotoxins. We also thank Dr. Aaron Darling for helpful discussions on phylogenetic analyses.

SUPPORTING INFORMATION AVAILABLE

αD-conopeptide cDNA sequences (Figure S1) and the amino acid and nucleotide changes in selected precursor pairs (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Lewis, R. J., and Garcia, M. L. (2003) Therapeutic potential of venom peptides. *Nat. Rev. Drug Discov. 2*, 790–802.
2. Nicke, A., Wonnacott, S., and Lewis, R. J. (2004) α-Conotoxins as tools for the elucidation of structure and function of neuronal nicotinic acetylcholine receptor subtypes. *Eur. J. Biochem. 271*, 2305–2319.
3. Loughnan, M., Nicke, A., Jones, A., Schroeder, C. I., Nevin, S. T., Adams, D. J., Alewood, P. F., and Lewis, R. J. (2006) Identification of a novel class of nicotinic receptor antagonists: Dimeric conotoxins VxXIIA, VxXIIB, and VxXIIC from *Conus vexillum*. *J. Biol. Chem. 281*, 24745–24755.
4. Jimenez, E. C., Olivera, B. M., and Teichert, R. W. (2007) αC-Conotoxin PrXA: A new family of nicotinic acetylcholine receptor antagonists. *Biochemistry 46*, 8717–8724.
5. Arias, H. R., and Blanton, M. P. (2000) Review: α-conotoxins. *Int. J. Biochem. Cell Biol. 32*, 1017–1028.
6. Terlau, H., and Olivera, B. M. (2004) *Conus* venoms: A rich source of novel ion channel-targeted peptides. *Physiol. Rev. 84*, 41–68.
7. Teichert, R. W., Jimenez, E. C., and Olivera, B. M. (2005) αS-Conotoxin RVIIIA: A structurally unique conotoxin that broadly targets nicotinic acetylcholine receptors. *Biochemistry 44*, 7897–7902.
8. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) Constant and hypervariable regions in conotoxin prepropeptides. *EMBO J. 9*, 1015–1020.
9. Sollod, B. L., Wilson, D., Zhaxybayeva, O., Gogarten, J. P., Drinkwater, R., and King, G. F. (2005) Were arachnids the first to use combinatorial peptide libraries? *Peptides 26*, 131–139.

10. 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) Speciation of cone snails and interspecific hyperdivergence of their venom peptides: Potential evolutionary significance of introns. *Ann. N.Y. Acad. Sci. 870*, 223–237.
11. Olivera, B. M. (2002) *Conus* venom peptides: Reflections from the biology of clades and species. *Annu. Rev. Ecol. Syst. 33*, 25–47.
12. Kauferstein, S., Melaun, C., and Mebs, D. (2005) Direct cDNA cloning of novel conopeptide precursors of the O-superfamily. *Peptides 26*, 361–367.
13. Kauferstein, S., Huys, I., Kuch, U., Melaun, C., Tytgat, J., and Mebs, D. (2004) Novel conopeptides of the I-superfamily occur in several clades of cone snails. *Toxicon 44*, 539–548.
14. Chen, W. H., Han, Y. H., Wang, Q., Miao, X. W., Ou, L., and Shao, X. X. (2006) cDNA cloning of two novel T-superfamily conotoxins from *Conus leopardus*. *Acta Biochim. Biophys. Sinica 38*, 287–291.
15. McIntosh, J. M., and Jones, R. M. (2001) Cone venom—from accidental stings to deliberate injection. *Toxicon 39*, 1447–1451.
16. Frohman, M. A. (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol. 218*, 340–356.
17. Griekspoor, A., and Groothuis, T. (1994) 4Peaks.
18. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. 10*, 1–6.
19. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol. 340*, 783–795.
20. Zhu, S., Bosmans, F., and Tytgat, J. (2004) Adaptive evolution of scorpion sodium channel toxins. *J. Mol. Evol. 58*, 145–153.
21. Dowell, C., Olivera, B. M., Garrett, J. E., Staheli, S. T., Watkins, M., Kuryatov, A., Yoshikami, D., Lindstrom, J. M., and McIntosh, J. M. (2003) α-Conotoxin PIA is selective for α6 subunit-containing nicotinic acetylcholine receptors. *J. Neurosci. 23*, 8445–8452.
22. McIntosh, J. M., Dowell, C., Watkins, M., Garrett, J. E., Yoshikami, D., and Olivera, B. M. (2002) α-Conotoxin GIC from *Conus geographus*, a novel peptide antagonist of nicotinic acetylcholine receptors. *J. Biol. Chem. 277*, 33610–33615.
23. LopezVera, E., Aguilar, M. B., Schiavon, E., Marinzi, C., Ortiz, E., Restano Cassulini, R., Batista, C. V., Possani, L. D., Heimer de la Cotera, E. P., Peri, F., Becerril, B., and Wanke, E. (2007) Novel α-conotoxins from *Conus spurius* and the α-conotoxin EI share high-affinity potentiation and low-affinity inhibition of nicotinic acetylcholine receptors. *FEBS J. 274*, 3972–3985.
24. Sandall, D. W., Satkunanathan, N., Keays, D. A., Polidano, M. A., Liping, X., Pham, V., Down, J. G., Khalil, Z., Livett, B. G., and Gayler, K. R. (2003) A novel α-conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves *in vivo*. *Biochemistry 42*, 6904–6911.
25. Ellison, M., Haberlandt, C., Gomez-Casati, M. E., Watkins, M., Elgoyhen, A. B., McIntosh, J. M., and Olivera, B. M. (2006) α-RgIA: A novel conotoxin that specifically and potently blocks the α9α10 nAChR. *Biochemistry 45*, 1511–1517.
26. Ellison, M., McIntosh, J. M., and Olivera, B. M. (2003) α-Conotoxins ImI and ImII. Similar α7 nicotinic receptor antagonists act at different sites. *J. Biol. Chem. 278*, 757–764.
27. Whiteaker, P., Christensen, S., Yoshikami, D., Dowell, C., Watkins, M., Gulyas, J., Rivier, J., Olivera, B. M., and McIntosh, J. M. (2007) Discovery, synthesis, and structure activity of a highly selective alpha7 nicotinic acetylcholine receptor antagonist. *Biochemistry 46*, 6628–6638.
28. Watkins, M., Olivera, B. M., Hillyard, D. R., McIntosh, J. M., and Jones, R. M. (2000)101-Conotoxin Peptides (Organization, W. I. P., Ed.) C07K 14/00, 14/435, 7/08, A61K 38/10, 38/17 ed., International.
29. Espiritu, D. J. D., Watkins, M., Dia-Monje, V., Cartier, G. E., Cruz, L. J., and Olivera, B. M. (2001) Venomous cone snails: molecular phylogeny and the generation of toxin diversity. *Toxicon 39*, 1899–1916.
30. Bulaj, G., DeLaCruz, R., Azimi-Zonooz, A., West, P., Watkins, M., Yoshikami, D., and Olivera, B. M. (2001) Delta-conotoxin structure/function though a cladistic analysis. *Biochemistry 40*, 13201–13208.
31. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.
32. Drummond, A. J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., and Wilson, A. (2007)Geneious v3.5 (available from http://www.geneious.com/).
33. Nei, M., and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol. 3*, 418–426.
34. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol. 24*, 1596–1599.
35. Brown, M. A., Begley, G. S., Czerwiec, E., Stenberg, L. M., Jacobs, M., Kalume, D. E., Roepstorff, P., Stenflo, J., Furie, B. C., and Furie, B. (2005) Precursors of novel Gla-containing conotoxins contain a carboxy-terminal recognition site that directs gamma-carboxylation. *Biochemistry 44*, 9150–9159.
36. Franco, A., Pisarewicz, K., Moller, C., Mora, D., Fields, G. B., and Mari, F. (2006) Hyperhydroxylation: a new strategy for neuronal targeting by venomous marine molluscs. *Prog. Mol. Subcell. Biol. 43*, 83–103.
37. Jakubowski, J. A., Keays, D. A., Kelley, W. P., Sandall, D. W., Bingham, J. P., Livett, B. G., Gayler, K. R., and Sweedler, J. V. (2004) Determining sequences and post-translational modifications of novel conotoxins in *Conus victoriae* using cDNA sequencing and mass spectrometry. *J. Mass Spectrom. 39*, 548–557.
38. Zhangsun, D., Luo, S., Wu, Y., Zhu, X., Hu, Y., and Xie, L. (2006) Novel O-superfamily conotoxins identified by cDNA cloning from three vermivorous *Conus* species. *Chem. Biol. Drug Des. 68*, 256–265.
39. Corpuz, G. P., Jacobsen, R. B., Jimenez, E. C., Watkins, M., Walker, C., Colledge, C., Garrett, J. E., McDougal, O., Li, W. Q., Gray, W. R., Hillyard, D. R., Rivier, J., McIntosh, J. M., Cruz, L. J., and Olivera, B. M. (2005) Definition of the M-conotoxin superfamily: Characterization of novel peptides from molluscivorous *Conus* venoms. *Biochemistry 44*, 8176–8186.
40. England, L. J., Imperial, J., Jacobsen, R., Craig, A. G., Gukyas, J., Akhtar, M., Rivier, J., Julius, D., and Olivera, B. M. (1998) Inactivation of a serotonin-gated ion channel by a polypeptide toxin from marine snails. *Science 281*, 575–578.
41. Jiang, H., Xu, C.-Q., Wang, C.-Z., Fan, C.-X., Zhao, T.-T., Chen, J.-S., and Chi, C.-W. (2006) Two novel O-superfamily conotoxins from *Conus vexillum*. *Toxicon 47*, 425–436.
42. Buczek, O., Yoshkami, D., Watkins, M., Bulaj, G., Jiminez, E. C., and Olivera, B. M. (2005) Characterization of D-amino-acid-containing excitatory conotoxins and redefinition of the I-conotoxin superfamily. *FEBS J. 272*, 4178–4188.
43. Wang, Q., Jiang, H., Han, Y.-H., Yuan, D.-D., and Chi, C.-W. (2008) Two different groups of signal sequence in M-superfamily conotoxins. *Toxicon 51*, 813–822.
44. Han, Y. H., Wang, Q., Jiang, H., Miao, X. W., Chen, J. S., and Chi, C. W. (2005) Sequence diversity of T-superfamily conotoxins from *Conus marmoreus*. *Toxicon 45*, 481–487.
45. Pi, C., Liu, J., Peng, C., Jiang, X., Zhao, Y., Tang, S., Wang, L., Dong, M., Chen, S., and Xu, A. (2006) Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics 88*, 809–819.
46. Seidah, N. G., Day, R., Marcinkiewicz, M., and Chretien, M. (1998) Precursor convertases: an evolutionary ancient, cell-specific, combinatorial mechanism yielding diverse biocative peptides and proteins. *Ann. N.Y. Acad. Sci. 839*, 9–24.
47. Buczek, O., Bulaj, G., and Olivera, B. M. (2005) Conotoxins and the posttranslational modification of secreted gene products. *Cell. Mol. Life Sci. 62*, 3067–3079.
48. Bandyopadhyay, P. K., Colledge, C. J., Walker, C. S., Zhou, L. M., Hillyard, D. R., and Olivera, B. M. (1998) Conantokin-G precursor and its role in gamma-carboxylation by a vitamin K-dependent carboxylase from a *Conus* snail. *J. Biol. Chem. 273*, 5447–5450.
49. Bush, K. A., Stenflo, J., Roth, D. A., Czerwiec, E., Harrist, A., Begley, G. S., Furie, B. C., and Furie, B. (1999) Hydrophobic amino acids define the carboxylation recognition site in the precursor of the gamma-carboxyglutamic-acid-containing conotoxin epsilon-TxIX from the marine cone snail *Conus textile*. *Biochemistry 38*, 14660–14666.
50. Grant, G. A., Al Rabiee, R., Xu, X. L., and Zhang, Y. P. (1997) Critical interactions at the dimer interface of kappa-bungarotoxin,

a neuronal nicotinic acetylcholine receptor antagonist. *Biochemistry 36*, 3353–3358.
51. Dewan, J. C., Grant, G. A., and Sacchettini, J. C. (1994) Crystal structure of kappa-bungarotoxin at 2.3 angstrom resolution. *Biochemistry 33*, 13147–13154.
52. Chiappinelli, V. A., and Wolf, K. M. (1989) Kappa-neurotoxins—Heterodimer formation between different neuronal nicotinic receptor antagonists. *Biochemistry 28*, 8543–8547.
53. Conticello, S. G., Pilpel, Y., Glusman, G., and Fainzilber, M. (2000) Position-specific codon conservation in hypervariable gene families. *Trends Genet. 16*, 57–59.
54. Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., and Fainzilber, M. (2001) Mechanisms for evolving hypervariability: The case of conopeptides. *Mol. Biol. Evol. 18*, 120–131.
55. Luo, S., Zhangsun, D., Zhang, B., Quan, Y., and Wu, Y. (2006) Novel α-conotoxins identified by gene sequencing from cone snails native to Hainan, and their sequence diversity. *J. Pept. Sci. 12*, 693–704.
56. Pi, C., Liu, Y., Peng, C., Jiang, X., Liu, J., Xu, B., Yu, X., Yu, Y., Jiang, X., Wang, L., Dong, M., Chen, S., and Xu, A.-L. (2006) Analysis of expressed sequence tags from the venom ducts of *Conus striatus:* focusing on the expression profile of conotoxins. *Biochimie 88*, 131–140.
57. Tennessen, J. A. (2005) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *J. Mol. Evol. 18*, 1387–1394.
58. Hughes, A. L. (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity 99*, 364–373.
59. Ogawa, T. (2006) Molecular diversity of proteins in biological offense and defense systems. *Mol. Divers. 10*, 511–514.
60. Yang, Z. H. (2005) The power of phylogenetic comparison in revealing protein function. *Proc. Natl. Acad. Sci. U.S.A. 102*, 3179–3180.
61. Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004) Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. *Mol. Biol. Evol. 21*, 1781–1791.
62. Mayrose, I., Doron-Faigenboim, A., Bacharach, E., and Pupko, T. (2007) Towards realistic codon models: among site variability and dependency of synonymous and non-synonymous rates. *Bioinformatics 23*, I319–I327.
63. Yang, Z. H., Nielsen, R., Goldman, N., and Pedersen, A. M. K. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics 155*, 431–449.
64. Tennessen, J. A. (2008) Positive selection drives a correlation between non-synonymous/synonymous divergence and functional divergence. *Bioinformatics 24*, 1421–1425.
65. Dewan, K. D. (2006) Secondary structure formations of conotoxin genes: A possible role in mediating variability. *Biochem. Biophys. Res. Commun. 349*, 701–708.
66. Olivera, B. M. (2006) *Conus* peptides: Biodiversity-based discovery and exogenomics. *J. Biol. Chem. 281*, 31173–31177.
67. Gray, W. R., Luque, A., Olivera, B. M., Barrett, J., and Cruz, L. J. (1981) Peptide toxins from *Conus geographus* venom. *J. Biol. Chem. 256*, 4734–4740.
68. Cruz, L. J., Gray, W. R., Olivera, B. M., Zeikus, R. D., Kerr, L., Yoshikami, D., and Moczydlowski, E. (1985) *Conus geographus* toxins that discriminate between neuronal and muscle sodium channels. *J. Biol. Chem. 260*, 9280–9288.
69. Hopkins, C., Grilley, M., Miller, C., Shon, K.-J., Cruz, L. J., Gray, W. R., Dykert, J., Rivier, J., Yoshikami, D., and Olivera, B. M. (1995) A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J. Biol. Chem. 270*, 22361–22367.
70. Walker, C. S., Steel, D., Jacobsen, R. B., Lirazan, M. B., Cruz, L. J., Hooper, D., Shetty, R., DelaCruz, R. C., Nielsen, J. S., Zhou, L. M., Bandyopadhyay, P., Craig, A. G., and Olivera, B. M. (1999) The T-superfamily of conotoxins. *J. Biol. Chem. 274*, 30664–30671.
71. Olivera, B. M., McIntosh, J. M., Cruz, L. J., Luque, F. A., and Gray, W. R. (1984) Purification and sequence of a presynaptic peptide toxin from *Conus geographus* venom. *Biochemistry 23*, 5087–5090.
72. Lirazan, M. B., Hooper, D., Corpuz, G. P., Ramilo, C. A., Bandyopadhyay, P., Cruz, L. J., and Olivera, B. M. (2000) The spasmodic peptide defines a new conotoxin superfamily. *Biochemistry 39*, 1583–1588.
73. 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) Novel excitatory *Conus* peptides define a new conotoxin superfamily. *J. Neurochem. 85*, 610–621.
74. Aguilar, M. B., LopezVera, E., Ortiz, E., Becerril, B., Possani, L. D., Olivera, B. M., and de la Cotera, E. P. H. (2005) A novel conotoxin from *Conus delessertii* with posttranslationally modified lysine residues. *Biochemistry 44*, 11130–11136.
75. Imperial, J. S., Bansal, P. S., Alewood, P. F., Daly, N. L., Craik, D. J., Sporning, A., Terlau, H., LopezVera, E., Bandyopadhyay, P. K., and Olivera, B. M. (2006) A novel conotoxin inhibitor of Kv1.6 channel and nAChR subtypes defines a new superfamily of conotoxins. *Biochemistry 45*, 8331–8340.
76. McIntosh, J. M., Corpuz, G. O., Layer, R. T., Garrett, J. E., Wagstaff, J. D., Bulaj, G., Vyazovkina, A., Yoshikami, D., Cruz, L. J., and Olivera, B. M. (2000) Isolation and characterization of a novel *Conus* peptide with apparent antinociceptive activity. *J. Biol. Chem. 275*, 32391–32397.
77. 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., and 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.

BI9000326