Published in final edited form as:

J Proteomics. 2015 January 30; 114: 38–47. doi:10.1016/j.jprot.2014.11.003.

Discovery by proteogenomics and characterization of an RF-amide neuropeptide from cone snail venom

Samuel D. Robinson^{a,1,*}, Helena Safavi-Hemami^{b,1,**}, Shrinivasan Raghuraman^b, Julita S. Imperial^b, Anthony T. Papenfuss^c, Russell W. Teichert^b, Anthony W. Purcell^d, Baldomero M. Olivera^b, and Raymond S. Norton^a

Samuel D. Robinson: sam.robinson@monash.edu; Helena Safavi-Hemami: safavihelena@gmail.com

^aMedicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville 3052, Australia

^bDepartment of Biology, University of Utah, Salt Lake City, Utah 84112, USA

^cBioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, VIC 3052, Australia

^dThe Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, VIC 3800, Australia

Abstract

In this study, a proteogenomic annotation strategy was used to identify a novel bioactive peptide from the venom of the predatory marine snail *Conus victoriae*. The peptide, conorfamide-Vc1 (CNF-Vc1), defines a new gene family. The encoded mature peptide was unusual for conotoxins in that it was cysteine-free and, despite low overall sequence similarity, contained two short motifs common to known neuropeptides/hormones. One of these was the C-terminal RF-amide motif, commonly observed in neuropeptides from a range of organisms, including humans. The mature venom peptide was synthesized and characterized structurally and functionally. The peptide was bioactive upon injection into mice, and calcium imaging of mouse dorsal root ganglion (DRG) cells revealed that the peptide elicits an increase in intracellular calcium levels in a subset of DRG neurons. Unusually for most *Conus* venom peptides, it also elicited an increase in intracellular calcium levels in a subset of non-neuronal cells.

Keywords

calcium imaging; conotoxin; neuropeptide; proteogenomics; RF-amide; venom

© 2014 Elsevier B.V. All rights reserved.

*Correspondence to: S. D. Robinson, Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville, VIC 3052, Australia, Tel.: +61 3 9903 9073. **Correspondence to: H. Safavi-Hemami, Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA, Tel.: +1 801 581 8370.

¹These authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Advances in high-throughput DNA sequencing have vastly improved the accessibility of genomic and transcriptomic information, but it is becoming clear that our current strategies for interpreting this huge volume of data are inadequate. In this post-genomic era, the focus is shifting to improving the assignment of biological function to a nucleic acid sequence. Current genome/transcriptome annotation methods, where programs like BLAST [1] are used to search a database of previously described sequences, rely on sequence similarity data and therefore suffer from an inability to annotate entirely novel genes/transcripts. Mass spectrometry (MS)-based proteomics, where experimentally-derived tandem MS (MS/MS) data are searched against a 6-frame translation of genomic or transcriptomic sequences [2], is emerging as a powerful annotation tool. This method is referred to as 'proteogenomics' [3] and has several advantages over solely genomic-based analyses: it demonstrates that a protein or peptide actually exists, establishes the correct protein/peptide sequence, identifies any post-translational modifications and, most importantly, can reveal novel genes/transcripts overlooked by traditional annotation methods (Figure 1A).

The goal of this work was to explore cone snail venom for novel classes of venom components. Cone snails (genus *Conus*) are carnivorous marine molluscs that use venom to incapacitate and capture their prey. *Conus* venoms are complex mixtures of hundreds of different peptides referred to as conotoxins. Individual conotoxins display often unique potency and selectivity profiles for a specific neuronal target, which can include specific subtypes of ion channels, G-protein coupled receptors (GPCRs) and neurotransmitter transporters [4–6]. As such, several conotoxins have found applications as research tools, some are being developed as drug leads and therapeutics, and there remains a concerted drive for the discovery of new conotoxins with novel bioactivity profiles.

We recently described the identification of over 100 conotoxin sequences in the venom gland transcriptome of *Conus victoriae* using a profile hidden Markov model (pHMM)-based annotation approach [7], but hypothesized that there must be sequences encoding bioactive peptides present in the venom gland transcriptome that remained undetected by pHMM and traditional annotation methods such as BLAST. Using a proteogenomic annotation strategy we identified a novel peptide precursor and its corresponding bioactive peptide, CNF-Vc1.

A number of unusual features make CNF-Vc1 unique among *Conus* venom components. Most notably, it has the C-terminal dipeptide Arg-Phe-NH₂ (RF-amide), which is the signature motif of the RF-amide family of neuropeptides. RF-amides are a diverse class of neuropeptides that share a C-terminal RF-amide motif (which can be broadened to include RX-amide where X is another hydrophobic residue), and are found in a range of organisms including humans where they exert diverse physiological effects mediated through specific GPCRs [8].

Here we describe the approach used to identify this new RF-amide peptide from cone snail venom, as well as its functional and structural characterization and subsequent examination of its unique activity profile.

2. Materials and methods

2.1. Venom gland transcriptome

Specimens of *C. victoriae* were collected from Broome, Western Australia. Snails were collected specifically for research use, under a commercial fishing license of the Western Australian Specimen Shell Managed Fishery (license number 2577). Ethics approval was not required, in Australia, for taking samples from *Conus*. Preparation of the venom gland transcriptome of *C. victoriae* has been described previously [7]. Briefly, whole venom glands of live specimens were dissected, snap-frozen in liquid nitrogen and stored at -80°C . Frozen venom glands were pulverized and homogenized prior to extraction of total RNA with TRIzol (Invitrogen, Life Technologies). cDNA library preparation, normalization and 454 sequencing were performed by Eurofins, MWG Operon (Budendorf, GER). *De novo* transcriptome assembly was performed using MIRA3 [9]. The assembled transcriptome was translated into 6-frames from which open reading frames longer than 40 residues were extracted and used as a database for subsequent MS-matching. The signal peptide sequence was determined using the SignalP 4.1 server [10].

2.2. Venom preparation

Venom was obtained by manual extrusion from freshly dissected venom glands, snap-frozen in liquid nitrogen and stored at -80°C . Extruded venom (from several specimens) was reconstituted in 0.1% TFA, pooled and homogenized using a glass Dounce tissue grinder. Insoluble material was pelleted by centrifugation, supernatant collected and lyophilized. Pellets were resuspended in 0.1% TFA / 20% acetonitrile (MeCN), then centrifuged and the supernatant was collected and lyophilized. This process was repeated with 40% and 60% MeCN. Lyophilized venom was resuspended in 2% MeCN, 0.1% TFA and pooled. Protein concentration was determined using a modified Bradford assay with ovalbumin as the standard. An aliquot of venom was reduced in 20 mM tris(2-carboxyethyl)phosphine (pH 8) for 30 min at 60°C . The sample was then alkylated by incubating in 40 mM iodoacetamide for 30 min. Lyophilized injected venom from *C. victoriae* was purchased from BioConus (www.bioconus.com). These specimens had also been sourced from Broome, WA, and maintained in captivity where injected venom was collected using a procedure adapted from Hopkins et al [11]. The injected venom sample was pooled from several individuals. Injected venom samples were prepared as described for extruded venom.

2.3. Mass spectrometry

Aliquots of 0.5 μg of each venom sample were centrifuged at $13,000 \times g$ for 10 min and the supernatant loaded onto a microfluidic trap column packed with ChromXP C18-CL 3 μm particles (300 $\text{\AA}$ nominal pore size; equilibrated in 0.1% formic acid/5 % MeCN) at 5 $\mu\text{L}/\text{min}$ using a Eksigent NanoUltra cHiPLC system. An analytical (15 cm $\times$ 75 μm ChromXP C18-CL 3) microfluidic column was then switched in line and peptides separated using linear gradient elution of 0–80 % MeCN/ 0.1 % formic acid over 90 min (300 nL/min). Separated peptides were analyzed using an AB SCIEX 5600 TripleTOF mass spectrometer equipped with a Nanospray III ion source and accumulating 30 MS/MS spectra per second.

2.4. Mass Spectrometry matching

Data were processed in ProteinPilot software (version Beta 4.1.46) using the Paragon algorithm. The search database comprised the *C. victoriae* venom gland transcriptome (described in experimental procedures). Peptides identified by ProteinPilot were then validated by comparison of experimentally-derived peaks against a theoretical peak list generated using Protein Prospector version 5.12.1 (prospector.ucsf.edu/prospector/mshome.htm).

2.5. Peptide synthesis and purification

The peptide was synthesized on an Apex 396 automated peptide synthesizer (AAPPTec, Louisville, KY) by applying standard solid-phase Fmoc (9-fluorenylmethyloxycarbonyl) protocols as previously described [12]. All standard amino acids were purchased from AAPPTec. The peptide was cleaved and purified as described previously [12]. Briefly, cleavage was performed from 30 mg of resin by a 4 h treatment with 0.5 mL of Reagent K (TFA/water/phenol/thioanisole/1,2-ethanedithiol 82.5/5/5/2.5 by volume) and subsequently filtered and precipitated with 12 mL of cold methyl-*tert*-butyl ether (MTBE). The crude peptide was collected by centrifugation and washed twice with 10 mL of cold MTBE. The washed peptide pellet was dissolved in 5 % (vol/vol) MeCN in 0.1% TFA in water and purified by reversed-phase high-performance liquid chromatography (RP-HPLC) using a semi-preparative C₁₈ column (Vydac, 218TP510, 250 mm × 10 mm, 5 µm particle size) and a gradient of 20 to 50 % solvent B over 30 min at a flow rate of 4 mL/min. HPLC solvents were 0.1% (vol/vol) TFA in water (solvent A) and 0.1% TFA (v/v) in 90% aqueous MeCN (solvent B). The absorbance of the eluent was monitored at 220/280 nm. Purity of the peptide was assessed by RP-HPLC using an analytical C₁₈ column (Vydac, 218TP54, 250 mm × 4.6 mm, 5 µm particle size) with a linear gradient of 20 to 50% solvent B over 30 min at a flow rate of 1 mL/min (Supplemental Figure S3). Peptide quantity was determined spectrophotometrically using the peptide's molar absorption coefficient (5690 cm⁻¹ M⁻¹). Molecular mass was confirmed by MALDI-TOF MS (Voyager, AB SCIEX) (Supplemental Figure S3).

2.6. NMR spectroscopy

The sample was prepared by dissolving the freeze-dried peptide in either 93% H₂O/7% ²H₂O, pH 3.5 or 100% C²H₃OH (Cambridge Isotope Laboratories) to a concentration of ~2.2 mM. One-dimensional ¹H spectra, two dimensional homonuclear TOCSY spectra (spin lock time 80 ms), DQF-COSY and NOESY spectra (mixing time 200 ms) were acquired for each sample condition at 293 K on a Bruker 600 MHz instrument. Chemical shift assignments for backbone and side chain protons were made by conventional analysis of two-dimensional TOCSY and NOESY spectra. Dioxane (chemical shift 3.75 ppm) and methanol (chemical shift 3.31 ppm) were used as internal references for the spectra acquired in water and methanol, respectively. All spectra were processed in TOPSPIN (version 3.2) and analysed using CcpNmr-Analysis (version 2.1.5).

2.7. Mouse bioassay

Swiss Webster mice (14–17 days old; 6.7 – 8.9 g) were injected intracranially (IC) with different doses of synthetic peptide dissolved in 12 μ L of 0.9% NaCl as described previously [13]. Screening was performed in duplicate at an initial dose of 20 nmol and repeated at lower concentrations until the lowest concentration showing a change in behavior was determined. Negative control mice were injected with 12 μ L of 0.9% NaCl. Following IC injection, mouse behavior was observed for 1 h to determine differences between treated and control animals.

2.8. Calcium imaging

Calcium imaging was carried out as described previously [14, 15]. Briefly, lumbar DRG neurons from 14–15 day old wild-type C57BL/6 mice were dissociated and plated on a 24-well poly-D-lysine-coated culture plate for the Ca^{2+} -imaging experiments described below. A total of four experiments was performed on cells from two mice. Cells were loaded with Fura-2 dye by addition of Fura-2-AM to their growth medium for 45 min at 37 °C, followed by 15 min at room temperature, after which the medium containing Fura-2-AM was replaced with observation solution (145 mM NaCl, 5 mM KCl, 2 mM CaCl_2 , 1 mM MgCl_2 , 1 mM sodium citrate, 10 mM HEPES, and 10 mM glucose) at room temperature for calcium imaging. Changes in cytosolic calcium concentration ($[\text{Ca}^{2+}]_i$) were monitored over time by standard ratiometric calcium imaging methods. Over the course of the experiment, at 7-min intervals, an elevated concentration of extracellular potassium ions (K^+) (25 mM) was applied to the cells for 15 s. This procedure resulted in a brief spike in $[\text{Ca}^{2+}]_i$ and is referred to as a K^+ pulse. Between the third and fourth K^+ pulses, CNF-Vc1 was applied for 6 min to the cells. A direct increase in $[\text{Ca}^{2+}]_i$ during the 6-min exposure to CNF-Vc1 and/or an indirect increase of the following K^+ pulse were considered a response to the peptide. Finally, to identify whether CNF-Vc1 targeted any specific subclass of neurons, the following compounds were applied after the sixth K^+ pulse: ATP (at min 43), menthol (at min 50) and capsaicin (at min 57). In each experiment, the calcium levels in the cell bodies of ~200 cells were monitored individually and simultaneously. All experiments involving the use of animals were approved by the Institutional Animal Care and Use Committee of the University of Utah.

2.9. Data availability

The nucleotide sequence of the CNF-Vc1 precursor has been deposited at DDBJ/EMBL/GenBank [Accession: GAIH01000072.1].

Chemical shift values for CNF-Vc1 in water and methanol have been deposited in the BMRB (ID: 25216).

3. Results

3.1. Discovery of a novel venom peptide

High-resolution MS/MS of the crude venom of *C. victoriae* generated 4,813 MS/MS spectra. This library of spectra was matched against the venom gland transcriptome database (the nature of which is described in the methods section) using ProteinPilot software.

Unsurprisingly, the vast majority of hits were of mature peptides derived from known conotoxin superfamilies (which will be described elsewhere). However, this combined approach also allowed the identification of an unusual *Conus* venom peptide derived from a peptide precursor that was distinct from any described previously.

The venom peptide was 18 residues in length, cysteine-free and had an amidated C-terminus (Figure 1B, Supplemental Figure S1, Supplemental Table S1). The full-length peptide had a C-terminal RF-amide motif, a motif that has been reported in neuropeptides in a range of organisms including humans [8]. Two RF-amide peptides have been isolated previously from *Conus spurius* venom [16, 17] and were called conorfamide-Sr1 and conorfamide-Sr2. We therefore, chose to name this peptide conorfamide-Vc1, abbreviated CNF-Vc1. It should be noted that, although CNF-Vc1 had the signature motif of the RF-amide family of neuropeptides, the amino-acid sequence was quite different from any other RF-amide peptide characterized previously.

The precursor of CNF-Vc1, predicted from the transcriptome data, was relatively short (59 residues) in comparison to that of most other conotoxins (Figure 1B). *Conus* venom peptide precursors typically have a three-domain arrangement, where the N-terminal signal sequence is followed by a long propeptide region and the mature peptide is encoded at the C-terminal end of the precursor [18]. In this regard, the precursor arrangement of CNF-Vc1, where the mature peptide is encoded between two propeptide regions, is atypical. The secretory signal peptide sequence (N-terminal 19 residues) of CNF-Vc1 was unique among conotoxins, sharing a maximum of only 66% sequence identity with that of the next most similar superfamily (O3). The N-terminal propeptide region is unusually short (6 residues), and ends in a dibasic (KK) cleavage site adjacent to the mature peptide. The C-terminal propeptide is slightly longer at 15 residues, and is also separated from the mature peptide region by a dibasic cleavage site (RR). A C-terminal Gly is removed from the mature peptide, yielding its amidated C-terminus. Investigation of the transcript encoding this sequence revealed that it was the sixth most highly expressed (according to assembled read count) in our normalized cDNA library. For example, the read count (2,294 assembled reads) was comparable to that of Vc1.1 (2,383 assembled reads), a potent analgesic and, until now, the only bioactive peptide from *C. victoriae* venom to be characterized [19]. This suggests a key role in the venom for this novel peptide.

To provide further support for the role of this peptide in envenomation and distinguish it from an endogenous neuropeptide, we repeated the method used above on a sample of injected *C. victoriae* venom (instead of venom obtained from dissected venom glands). Our injected venom sample was less complex, producing a more simple LC profile and generating 3,062 MS/MS spectra. A search of this library identified the MH^{+4} precursor ion and corresponding MS/MS spectra identical to that observed for CNF-Vc1 in the extruded venom (Supplemental Table S1), confirming the presence of this peptide in injected venom. These results unequivocally demonstrate that CNF-Vc1 functions as an exogenously targeted venom component in *C. victoriae*.

In addition to CNF-Vc1, several other novel peptides were also identified using this proteogenomic strategy (Supplemental Table S1). Although overlooked by similarity-based

annotation, these peptides were derived from ‘conotoxin-like’ precursors, each with a secretory signal peptide, followed by a propeptide region and the mature peptide encoded at the C-terminus (Supplemental Figure S2). Two of the encoded mature peptides were cysteine-rich, each with a type VI/VII cysteine framework [20], common among conotoxins. The remaining peptides were cysteine-free and all derived from a single unusual precursor. Given their presence in the venom and similarity to known conotoxins it seemed reasonable to assume that these sequences represented a new class of conotoxin. Indeed, similar peptides have since been described in the venom gland transcriptome of *Conus marmoreus* [21], where they were designated as a new conotoxin gene superfamily, the H-superfamily. It will be of interest to explore the functional properties of this new class of conotoxins, but the rest of this study focuses on the functional characterization of CNF-Vc1.

In order to obtain sufficient quantity of CNF-Vc1 for a comprehensive functional and structural characterization, the peptide was synthesized by standard solid-phase peptide synthesis (Supplemental Figure S3). A comparison of the MS/MS spectra of native and synthetic CNF-Vc1 is shown in Supplemental Figure S4. The same fragmentation pattern is observed for both the synthetic and native peptide.

3.2 NMR spectroscopy

Initial examination of one-dimensional ^1H NMR spectra (Supplemental Figure S5) indicated that the peptide existed as a single major species in solution, with some evidence of a minor species under both sample conditions (water and methanol). Poor peak dispersion, in both water and methanol, indicated that the peptide did not adopt ordered secondary structure in solution.

Strong $d_{\alpha\delta}$ NOEs for Gly10-Pro11 were indicative of a *trans* conformation about the X-Pro bond. The H^α resonance for His1 was identified by a Ser2 H^N – His1 H^α NOESY cross-peak and subsequent identification of the H^α - H^β cross-peak. ^1H chemical shift assignments are tabulated in Supplemental Tables S2 and S3. Comparison of H^α chemical shifts to random coil values [22] confirmed a lack of ordered secondary structure for CNF-Vc1 in both water and methanol (Supplemental Figure S6).

3.3 Behavioral assay

IC injection of the peptide in mice elicited a strong behavioral response (Table 1); immediately following injection (10 and 20 nmol) the mice were completely incapacitated, with splayed legs and no movement other than leg and head twitches ($n = 3$ mice). The recovery phase started at 45 min with mice starting to walk around in the cage. However, body muscle twitching remained until the end of the experiment (1 h post-injection). The lowest dose causing this phenotype was 2.5 nmol ($n = 2$ mice). Interestingly, mice showed signs of hyperactivity at this dose, that remained down to a dose of 0.5 nmol ($n = 2$ mice). When the dose was lowered further to 0.3 nmol, mice showed no difference in behavior from control mice ($n = 1$ mouse).

3.4 Calcium imaging assay

In order to further evaluate the basis for the bioactivity observed *in vivo*, the effects of the peptide were assessed using calcium imaging. The specific experimental paradigm used a primary culture of dissociated mouse dorsal-root ganglion (DRG) cells loaded with a calcium sensitive dye (Fura-2) to simultaneously monitor the responses of individual cells in a well containing ~200 cells [23]. As the responses of individual cells can be monitored, the effects of a bioactive peptide can be evaluated in any and all of the neuronal and glial subclasses present in the DRG. Results obtained by applying CNF-Vc1 to a culture of DRG cells are shown in Figure 2.

A significant fraction of all DRG neurons responded to the peptide with increases in cytosolic calcium elicited in the affected cells. Most small- and medium-sized neurons were *not* affected by the peptide, indicating that the peptide does not act indiscriminately on all neurons. However, the majority of large-diameter DRG neurons was affected. Responses were typically characterized by a delayed (2–5 min following application) but sudden increase in $[Ca^{2+}]_i$, which fluctuated in most neurons. This was typically accompanied by a potentiation of the following depolarization elicited by a high concentration of extracellular K^+ . In most cases responses were reversible over the course of the experiment. Several traces illustrating individual neuronal responses are shown in Figure 2B. Certain neuronal subclasses were unresponsive to CNF-Vc1, as shown in Supplemental Figure S7; these included neurons that responded to menthol, which comprise cold thermosensors and cold nociceptors [24].

One particularly notable feature of CNF-Vc1 is that it elicited an increase in $[Ca^{2+}]_i$ in a large fraction of non-neuronal cells. Most peptides from *Conus* venoms act only on a subset of neurons and do not affect non-neuronal cells (our unpublished observations). The typical response to CNF-Vc1 in non-neuronal DRG cells was characterized by a delayed but sudden increase in $[Ca^{2+}]_i$, which remained static until washout (Figure 2C).

In order to determine the minimum effective concentration of CNF-Vc1, the peptide was applied in increasing concentrations until an effect was observed. The minimum effective concentration of CNF-Vc1 was 5 μ M (Supplemental Figure S8).

Together, the results demonstrate that this peptide acts selectively on a diverse spectrum of neuronal subclasses, and a substantial fraction of non-neuronal cells, (presumably glia) present in the DRG. In contrast, another RF-amide peptide (conorfamide-Sr1 (CNF-Sr1) from *C. spurius* venom) did not affect DRG cells; CNF-Sr1 was compared directly to CNF-Vc1 in the experiment shown in Figure 3. While CNF-Vc1 had effects on both neuronal and non-neuronal cells, these cells were not affected by CNF-Sr1.

4. Discussion

This study documents our use of a proteogenomics strategy to identify a novel bioactive peptide in the venom of *C. victoriae*. The high level of expression of this peptide, its presence in both extruded and injected venom, and its strong bioactivity profile demonstrate that this peptide is likely to play a key role in the envenomation strategy of *C. victoriae*.

However, given its limited sequence identity with other genes or gene products, traditional annotation methods were not able to annotate this peptide as such. This highlights the power of proteogenomics as a strategy to annotate previously undefined genes/transcripts, as recognized and first employed by Yates et al [2]. As noted previously [3], a more widespread application of this annotation strategy offers the opportunity to unlock currently hidden regions of genomes/transcriptomes.

Conforamide_Vc1 defines a new *Conus* venom peptide gene family. All conotoxins are post-translationally processed from larger precursors and each conotoxin precursor can be grouped into a gene superfamily based on signal sequence [18]. The signal sequence of CNF-Vc1 shares some similarity with that of the O3-superfamily conotoxins (Figure 1B), but the remainder of the precursor is quite distinct. An unusual feature of the CNF-Vc1 precursor is its propeptide. The typical *Conus* venom peptide precursor, including those of the O3-superfamily, has an N-terminal signal sequence followed by a propeptide region, which generally varies between 20 and 40 residues in length, and finally the mature peptide at the C-terminus [18]. The CNF-Vc1 precursor is unusual in that the propeptide region between the signal sequence and the mature peptide toxin is extremely short (6 residues), and there is a second propeptide region at the C-terminus, excised to generate the mature peptide identified from venom. Thus, processing of the precursor to the mature venom peptide is clearly unusual compared to most *Conus* venom peptides. Furthermore, all O3-superfamily precursors characterized so far encode a cysteine-rich mature peptide with the type VI/VII cysteine framework (C...C...CC...C...C) whereas CNF-Vc1 lacks any cysteine residues. These features distinguish CNF-Vc1 from the O3-superfamily. Nevertheless, the similarity with the O3-superfamily signal sequence suggests a distant but definite relationship between O3 venom peptides and CNF-Vc1.

Among its unusual features, CNF-Vc1 has the signature motif of the RF-amide family of neuropeptides (i.e., the C-terminal RF-amide dipeptide). In molluscs, RF-amide neuropeptides are widely expressed in the central and peripheral nervous systems where they regulate cardiovascular, reproductive and sensory systems, feeding and respiration [8]. In mammals RF-amide distribution patterns appear to be limited to specific brain areas where they are involved in essential functions such as cardiovascular regulation, pain modulation and food consumption. To date, RF-amides from five precursors have been characterized in humans [25]: neuropeptides FF (NPFF) and AF (NPAF) [26], Prolactin-releasing peptide (PrRP) [27], RFRPs 1 and 3 [28], 26RFa [29], and the kisspeptins [30]. Most of the neuronal actions of RF-amides appear to be mediated through specific GPCRs [27, 28, 30–33], although two ionotropic receptors for these peptides have been reported [34, 35], viz the invertebrate FMRF-amide-gated Na⁺ channel (FaNaC) and the related mammalian acid-sensing ion channels (ASICs).

While we grouped CNF-Vc1 with the RF-amide family of peptides, based on its C-terminal RF-amide motif, it is noteworthy that the N-terminal dipeptide His-Ser of CNF-Vc1 is reminiscent of another class of hormones/neuropeptides, which belong to the pituitary adenylate cyclase-activating polypeptide (PACAP)/ glucagon superfamily (Figure 4). The PACAP/ glucagon superfamily [36] includes nine hormones/neuropeptides in humans with a wide distribution, particularly in the brain and the gut, and each targets a specific GPCR.

This family of peptide hormones also includes the exendins, which were discovered in lizard venom [37–39].

The presence in *Conus* venom of an RF-amide peptide is not unprecedented. Two RF-amide peptides have previously been isolated from the extruded venom of *C. spurius*, collected off the Eastern coast of Mexico [16, 17]. *C. spurius* is a worm-hunting cone snail; in contrast, *C. victoriae*, collected off Western Australia, belongs to a snail-hunting clade (subgenus *Cylindrus*) of *Conus* species. It remains to be established how widely distributed similar peptides are in venoms of other species of cone snails. Other than its C-terminal RF-amide motif, CNF-Vc1 shares little sequence similarity with CNF-Sr1 and CNF-Sr2 from *C. spurius*. Both CNF-Sr1 and CNF-Sr2 displayed activity in the mouse central nervous system and the latter was also active on molluscan muscle [16, 17]. The hyperactivity induced by lower doses of CNF-Vc1 is consistent with what is reported for the *C. spurius* peptides. However, at higher doses (2.5–20 nmol) the responses recorded here for CNF-Vc1 appear distinct from, and more severe than, those reported for either of the *C. spurius* peptides, with complete incapacitation of the injected mouse lasting up to 20 min post-injection.

Perhaps the most significant feature of CNF-Vc1 is its unique activity profile in DRG cells. It causes an increase in cytosolic calcium in a specific subset of DRG neurons, and, in contrast to other *Conus* venom peptides that have been tested on DRG cell cultures (our unpublished observations), the peptide also affects a large fraction of the non-neuronal cells. CNF-Vc1 causes fluctuations in $[Ca^{2+}]_i$ in a range of neuronal and non-neuronal DRG cells. In contrast, CNF-Sr1 from *C. spurius* has no apparent activity on DRG cells (Figure 3). The different bioactivity of CNF-Vc1 from other RF-amide peptides raises the possibility that the peptide has molecular targets not affected by other RF-amide peptides. Thus, it is reasonable that the unique sequence features of CNF-Vc1, including the exendin-related features, could be responsible for some of the unusual biological effects of this peptide.

Our results highlight the power of a proteogenomics annotation strategy for the discovery of novel, biologically relevant transcripts and their products. The peptide discovered by this means, CNF-Vc1, is a potential new tool for clarifying the physiological role of the RF-amide and exendin-type neuropeptide families in both molluscan and mammalian nervous systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Johan Pas for specimen collection, Dorothy Loo and Dr. Nicholas Williamson for technical assistance with mass spectrometry, Dr. Joanna Gajewiak for peptide synthesis and My Thi Thao Huynh for assistance with image presentation. The authors acknowledge financial support from a Discovery Grant (DP110101331) from the Australian Research Council (BMO, AWP) and a National Institutes of Health Grant GM 48677 (BMO). AWP and RSN acknowledge fellowship support from the Australian National Health and Medical Research Council. HSH is supported by a Marie Curie Fellowship from the European Commission (CONBIOS 330486).

Abbreviations

ASIC	acid-sensing ion channel
CNF	conoramide
C-terminus	carboxyl terminus
[Ca²⁺]_i	cytosolic calcium ion concentration
DRG	dorsal root ganglion
FaNaC	FMRF-amide-gated Na ⁺ channel
GPCR	G protein-coupled receptor
IC	intracranial
MeCN	acetonitrile
MTBE	methyl- <i>tert</i> -butyl ether
N-terminus	amino terminus
PACAP	pituitary adenylate cyclase-activating polypeptide
pHMM	profile hidden Markov model

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–10. [PubMed: 2231712]
2. Yates JR, Eng JK, McCormack AL. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem.* 1995; 67:3202–10. [PubMed: 8686885]
3. Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD. Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Brief Funct Genomic Proteomic.* 2008; 7:50–62. [PubMed: 18334489]
4. Terlau H, Olivera BM. *Conus* venoms: A rich source of novel ion channel-targeted peptides. *Physiol Rev.* 2004; 84:41–68. [PubMed: 14715910]
5. Lewis RJ, Dutertre S, Vetter I, Christie MJ. *Conus* venom peptide pharmacology. *Pharmacol Rev.* 2012; 64:259–98. [PubMed: 22407615]
6. Norton RS, Olivera BM. Conotoxins down under. *Toxicon.* 2006; 48:780–98. [PubMed: 16952384]
7. Robinson SD, Safavi-Hemami H, McIntosh LD, Purcell AW, Norton RS, Papenfuss AT. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS ONE.* 2014; 9:e87648. [PubMed: 24505301]
8. Walker R, Papaioannou S, Holden-Dye L. A review of FMRFamide- and RFamide-like peptides in metazoa. *Invertebr Neurosci.* 2009; 9:111–53.
9. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, et al. Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Res.* 2004; 14:1147–59. [PubMed: 15140833]
10. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011; 8:785–6. [PubMed: 21959131]
11. Hopkins C, Grilley M, Miller C, Shon K-J, Cruz LJ, Gray WR, et al. A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J Biol Chem.* 1995; 270:22361–7. [PubMed: 7673220]

12. Zhang MM, Wilson MJ, Gajewiak J, Rivier JE, Bulaj G, Olivera BM, et al. Pharmacological fractionation of tetrodotoxin-sensitive sodium currents in rat dorsal root ganglion neurons by μ -conotoxins. *British Journal of Pharmacology*. 2013; 169:102–14. [PubMed: 23351163]
13. McIntosh JM, Yoshikami D, Mahe E, Nielsen DB, Rivier JE, Gray WR, et al. A nicotinic acetylcholine receptor ligand of unique specificity, alpha-conotoxin ImI. *Journal of Biological Chemistry*. 1994; 269:16733–9. [PubMed: 8206995]
14. Raghuraman S, Garcia AJ, Anderson TM, Twede VD, Curtice KJ, Chase K, et al. Defining modulatory inputs into CNS neuronal subclasses by functional pharmacological profiling. *Proc Natl Acad Sci USA*. 2014; 111:6449–54. [PubMed: 24733934]
15. Teichert RW, Smith NJ, Raghuraman S, Yoshikami D, Light AR, Olivera BM. Functional profiling of neurons through cellular neuropharmacology. *Proc Natl Acad Sci USA*. 2012
16. Maillou M, Aguilar MB, López-Vera E, Craig AG, Bulaj G, Olivera BM, et al. Conorfamide, a *Conus* venom peptide belonging to the RFamide family of neuropeptides. *Toxicon*. 2002; 40:401–7. [PubMed: 11738233]
17. Aguilar MB, Luna-Ramírez KS, Echeverría D, Falcón A, Olivera BM, Heimer de la Cotera EP, et al. Conorfamide-Sr2, a gamma-carboxyglutamate-containing FMRFamide-related peptide from the venom of *Conus spurius* with activity in mice and mollusks. *Peptides*. 2008; 29:186–95. [PubMed: 18201803]
18. Woodward SR, Cruz LJ, Olivera BM, Hillyard DR. Constant and hypervariable regions in conotoxin propeptides. *The EMBO journal*. 1990; 9:1015–20. [PubMed: 1691090]
19. Sandall DW, Satkunathan N, Keays DA, Polidano MA, Liping X, Pham V, et al. A novel α -conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves in vivo. *Biochemistry*. 2003; 42:6904–11. [PubMed: 12779345]
20. Kaas Q, Westermann JC, Craik DJ. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon*. 2010; 55:1491–509. [PubMed: 20211197]
21. Dutertre S, Jin A-h, Kaas Q, Jones A, Alewood PF, Lewis RJ. Deep venomomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol Cell Proteomics*. 2012
22. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids: Investigations of nearest-neighbor effects. *J Biomol NMR*. 1995; 5:67–81. [PubMed: 7881273]
23. Teichert RW, Raghuraman S, Memon T, Cox JL, Foulkes T, Rivier JE, et al. Characterization of two neuronal subclasses through constellation pharmacology. *Proc Natl Acad Sci USA*. 2012; 109:12758–63. [PubMed: 22778416]
24. Teichert RW, Memon T, Aman JW, Olivera BM. Using constellation pharmacology to define comprehensively a somatosensory neuronal subclass. *Proc Natl Acad Sci USA*. 2014; 111:2319–24. [PubMed: 24469798]
25. Findeisen M, Rathmann D, Beck-Sickinger AG. RFamide peptides: Structure, function, mechanisms and pharmaceutical potential. *Pharmaceuticals*. 2011; 4:1248–80.
26. Perry SJ, Yi-Kung Huang E, Cronk D, Bagust J, Sharma R, Walker RJ, et al. A human gene encoding morphine modulating peptides related to NPFF and FMRFamide. *FEBS Lett*. 1997; 409:426–30. [PubMed: 9224703]
27. Hinuma S, Habata Y, Fujii R, Yuji Kawamata, Hosoya M, Fukusumi S, et al. A prolactin-releasing peptide in the brain. *Nature*. 1998; 393:272–6. [PubMed: 9607765]
28. Hinuma S, Shintani Y, Fukusumi S, Iijima N, Matsumoto Y, Hosoya M, et al. New neuropeptides containing carboxy-terminal RFamide and their receptor in mammals. *Nat Cell Biol*. 2000; 2:703–8. [PubMed: 11025660]
29. Chartrel N, Dujardin C, Anouar Y, Leprince J, Decker A, Clerens S, et al. Identification of 26RFa, a hypothalamic neuropeptide of the RFamide peptide family with orexigenic activity. *Proc Natl Acad Sci USA*. 2003; 100:15247–52. [PubMed: 14657341]
30. Kotani M, Detheux M, Vandenbogaerde A, Communi D, Vanderwinden J-M, Le Poul E, et al. The metastasis suppressor gene KiSS-1 encodes kisspeptins, the natural ligands of the orphan G protein-coupled receptor GPR54. *J Biol Chem*. 2001; 276:34631–6. [PubMed: 11457843]

31. Bonini JA, Jones KA, Adham N, Forray C, Artymyshyn R, Durkin MM, et al. Identification and characterization of two G protein-coupled receptors for neuropeptide FF. *J Biol Chem.* 2000; 275:39324–31. [PubMed: 11024015]
32. Elshourbagy NA, Ames RS, Fitzgerald LR, Foley JJ, Chambers JK, Szekeres PG, et al. Receptor for the pain modulatory neuropeptides FF and AF is an orphan G protein-coupled receptor. *J Biol Chem.* 2000; 275:25965–71. [PubMed: 10851242]
33. Jiang Y, Luo L, Gustafson EL, Yadav D, Laverty M, Murgolo N, et al. Identification and characterization of a novel RF-amide peptide ligand for orphan G-protein-coupled receptor SP9155. *J Biol Chem.* 2003; 278:27652–7. [PubMed: 12714592]
34. Askwith CC, Cheng C, Ikuma M, Benson C, Price MP, Welsh MJ. Neuropeptide FF and FMRFamide potentiate acid-evoked currents from sensory neurons and proton-gated DEG/ENaC channels. *Neuron.* 2000; 26:133–41. [PubMed: 10798398]
35. Lingueglia E, Champigny G, Lazdunski M, Barbry P. Cloning of the amiloride-sensitive FMRFamide peptide-gated sodium channel. *Nature.* 1995; 378:730–3. [PubMed: 7501021]
36. Sherwood NM, Krueckl SL, McRory JE. The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr Rev.* 2000; 21:619–70. [PubMed: 11133067]
37. Parker DS, Raufman JP, O'Donohue TL, Bledsoe M, Yoshida H, Pisano JJ. Amino acid sequences of helospectins, new members of the glucagon superfamily, found in Gila monster venom. *J Biol Chem.* 1984; 259:11751–5. [PubMed: 6207171]
38. Eng J, Kleinman WA, Singh L, Singh G, Raufman JP. Isolation and characterization of exendin-4, an exendin-3 analogue, from *Heloderma suspectum* venom. Further evidence for an exendin receptor on dispersed acini from guinea pig pancreas. *J Biol Chem.* 1992; 267:7402–5. [PubMed: 1313797]
39. Eng J, Andrews PC, Kleinman WA, Singh L, Raufman JP. Purification and structure of exendin-3, a new pancreatic secretagogue isolated from *Heloderma horridum* venom. *J Biol Chem.* 1990; 265:20259–62. [PubMed: 1700785]

Highlights

- A proteogenomics-based annotation strategy was used to identify a new peptide.
- This peptide, CNF-Vc1, belongs to the RF-amide neuropeptide family.
- CNF-Vc1 has a unique activity profile in neuronal and non-neuronal cells.
- Proteogenomics is a powerful tool for the discovery of novel bioactive peptides.

Significance

Our findings illustrate the utility of proteogenomics for the discovery of novel, functionally relevant genes and their products. CNF-Vc1 should be useful for understanding the physiological role of RF-amide peptides in the molluscan and mammalian nervous systems.

Figure 1. Discovery of novel peptides by proteogenomics

(A) The tissue of interest, in this case the venom gland of *Conus victoriae* is first examined separately by genomic and proteomic methods to generate a genome/transcriptome and proteome respectively. These data are then combined via MS-matching which involves searching the proteome database, in the form of LC-MS/MS data, against a 6-frame translation of the corresponding genome/transcriptome. This strategy allows one to validate or correct previously predicted gene products, to establish the correct protein/peptide sequence and importantly has the power to identify truly novel genes and their products.

(B) The primary structure of the CNF-Vc1 venom peptide and corresponding precursor sequence. Precursor sequences of the two O3-superfamily conotoxins, the bromosleeper peptide (GenBank: GQ981406.1) and Ar6.25 (UniProt: Q9BP66), are shown for comparison. Precursor signal peptides are highlighted in purple, dibasic cleavage sites in red and encoded mature peptides in blue.

Figure 2. Application of CNF-Vc1 elicited fluctuating increases in $[Ca^{2+}]_i$ in neuronal and non-neuronal DRG cells

(A) Responses of DRG cells to CNF-Vc1. Top left, fluorescence associated with basal resting $[Ca^{2+}]_i$. Bottom left, cellular response to depolarization induced by a 15 s pulse of high extracellular K⁺ (25 mM), characterized by a large increase in $[Ca^{2+}]_i$ in viable DRG neurons. Top right, cellular response to application of CNF-Vc1 (15 μM), characterized by fluctuating increases in $[Ca^{2+}]_i$ in a subset of both neuronal and non-neuronal cells. Bottom

right, cellular response to depolarization induced by K^+ (25 mM), following the incubation of CNF-Vc1 (scale bar for relative $[Ca^{2+}]_i$ shown at right-hand side of panel).

(B) Representative calcium-imaging traces (representing changes in $[Ca^{2+}]_i$) from individual DRG neuronal cells. Each trace is the response of a single neuron from the same experimental trial. The top trace is an example of a response from a neuronal cell that remained unaffected by the incubation of CNF-Vc1. The bottom 4 traces are examples of 4 different neurons that responded to incubation of peptide with transient or fluctuating increases in $[Ca^{2+}]_i$. Cross-sectional cell areas are indicated to the right of the traces.

(C) Representative calcium-imaging traces from individual DRG non-neuronal cells. Each trace is the response of a single non-neuronal cell, as identified by the absence of an increase in $[Ca^{2+}]_i$ elicited by depolarization. Top trace is an example of a non-neuronal cell that did not respond to CNF-Vc1 incubation but responded to the application of ATP. Bottom 3 traces are examples of 3 non-neuronal cells that responded to both CNF-Vc1 incubation and to ATP application with increases in $[Ca^{2+}]_i$. Cross-sectional cell areas are indicated to the right of each trace.

The following applies to each calcium-imaging trace shown in B and C. Each arrow indicates the application of a compound to the bath solution for a duration of 15 s.

Abbreviations below each arrow indicate the compound applied at each time point.

Abbreviations are as follows: K, 25 mM K^+ (applied at 7 min intervals); ATP, 20 μ M (at min 43); M; 400 μ M menthol (at min 50); and C, 300 nM capsaicin (at min 57). The black horizontal bar indicates when cells were exposed to CNF-Vc1 in the bath solution. The y-axis is a relative measure of $[Ca^{2+}]_i$, determined by the 340/380 nm excitation ratio. The x-axis is the same for all traces (time in min).

Figure 3. Application of CNF-Vc1 elicited fluctuating increases in $[Ca^{2+}]_i$ in neuronal and non-neuronal DRG cells, in contrast to RF-amide peptide, CNF-Sr1, from *Conus spurius*

(A) Representative calcium-imaging traces (representing changes in $[Ca^{2+}]_i$) from individual DRG neuronal cells. Top two traces are from neuronal cells that were apparently unaffected by the incubation of CNF-Sr1 and CNF-Vc1. Bottom two traces are from neuronal cells that responded to CNF-Vc1 with increases in $[Ca^{2+}]_i$ but did not respond to the application of CNF-Sr1 in this manner.

(B) Representative calcium-imaging traces from individual DRG non-neuronal cells. Top trace is from a non-neuronal cell that responded to the application of ATP with an increase in $[Ca^{2+}]_i$ but did not respond to CNF-Vc1 in this manner. Bottom trace is from a non-neuronal cell that responded to both ATP and CNF-Vc1 with increases in $[Ca^{2+}]_i$, but did not respond to CNF-Sr1 in this manner. The following applies to all calcium-imaging traces in this figure. Each arrow indicates the application of a compound for a duration of 15 s. Abbreviations are as follows: K, 25 mM K^+ (applied at 7 min intervals); ATP, 20 μ M (at min 43); M, 400 μ M menthol (at min 50); and C, 300 nM capsaicin (at min 57). The horizontal grey bar indicates when cells were exposed to 15 μ M CNF-Sr1 in the bath

solution. The black horizontal bar indicates when the cells were exposed to 15 μ M CNF-Vc1 in the bath solution. The y-axis is a relative measure of $[Ca^{2+}]_i$, determined by the 340/380 nm excitation ratio. The x-axis is the same for all traces (time in min).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

CNF-Vc1	HS GFLLAWSGPRNRFV RF *
RF-amide neuropeptides	
FMRF-amide	FM RF *
NPFF	FLFQPQ RF *
CNF-Sr1	GPMGWVPVFY RF *
CNF-Sr2	GPM γ DPL γ I RI *
PACAP/Glucagon neuropeptides/hormones	
PACAP	HS DGIFTDSYSRYRKQMAVKKYLA AVL*
secretin	HS DGTFTSELSRLREGARLQRL LQGLV*
exendin	HS DATFTA EYSKLLAKLALQKYLE SILGSSTSPRPPSS

Figure 4. The venom peptide CNF-Vc1 shares structural similarity with RF-amide neuropeptides and peptide hormones of the PACAP/glucagon superfamily
 CNF-Vc1 is compared with selected RF-amide neuropeptides; FMRF-amide (*Lymnaea stagnalis*), neuropeptide-FF (*H. sapiens*), CNF-Sr1 (*C. spurius*) and CNF-Sr2 (*C. spurius*), and peptide hormones of the PACAP/glucagon superfamily; PACAP (*H. sapiens*), secretin (*H. sapiens*) and exendin-1 (*Heloderma suspectum*). Shared dipeptide motifs are in bold; *, C-terminal amidation; γ , gamma-carboxy glutamate.

Table 1

Summary of results from intracranial mouse injections. Mice were between 14 – 17 days old and weighed between 6.7 – 8.9 g. Injections were carried out in duplicate unless indicated otherwise.

Dose (nmol)	Observed Behaviour (time = approximate min post injection)
20	Immediate incapacitation after injection, 0–20 min: no movement other than hind leg twitching with splayed legs, 20–35 min: head upright but no other movement, head shaking uncontrollably, 35–45 min: moving around with occasional body twitching, 45–60 min: normal behaviour other than intermittent shaking of body
10 (n=1)	No difference from 20 nmol dose: incapacitation and twitching of hind legs and head, back to normal after 40 min
2.5	0–5 min: incapacitation with twitching limbs, recovery after 5–10 min followed by hyperactivity : climbing and jumping at high frequency, no resting until 60 min
0.5	0–1 min: stiff tail with no other signs of paralysis, 1–10 min: normal behaviour, 15–30 min: hyperactive : climbing and jumping at high frequency, 30–60 min: normal behaviour
0.3 (n=1)	No difference from control mice