

Combined Proteomic and Transcriptomic Interrogation of the Venom Gland of *Conus geographus* Uncovers Novel Components and Functional Compartmentalization*[§]

Helena Safavi-Hemami[‡], Hao Hu[§], Dhana G. Gorasia[¶], Pradip K. Bandyopadhyay[‡], Paul D. Veith[¶], Neil D. Young^{||}, Eric C. Reynolds[¶], Mark Yandell[§], Baldomero M. Olivera[‡], and Anthony W. Purcell^{**}

Cone snails are highly successful marine predators that use complex venoms to capture prey. At any given time, hundreds of toxins (conotoxins) are synthesized in the secretory epithelial cells of the venom gland, a long and convoluted organ that can measure 4 times the length of the snail's body. In recent years a number of studies have begun to unveil the transcriptomic, proteomic and peptidomic complexity of the venom and venom glands of a number of cone snail species. By using a combination of DIGE, bottom-up proteomics and next-generation transcriptome sequencing the present study identifies proteins involved in envenomation and conotoxin maturation, significantly extending the repertoire of known (poly)peptides expressed in the venom gland of these remarkable animals. We interrogate the molecular and proteomic composition of different sections of the venom glands of 3 specimens of the fish hunter *Conus geographus* and demonstrate regional variations in gene expression and protein abundance. DIGE analysis identified 1204 gel spots of which 157 showed significant regional differences in abundance as determined by biological variation analysis. Proteomic interrogation identified 342 unique proteins including those that exhibited greatest fold change. The majority of these proteins also exhibited significant changes in their mRNA expression levels validating the reliability of the experimental approach. Transcriptome sequencing further revealed a yet unknown genetic

diversity of several venom gland components. Interestingly, abundant proteins that potentially form part of the injected venom mixture, such as echotoxins, phospholipase A₂ and con-ikots-ikots, classified into distinct expression clusters with expression peaking in different parts of the gland. Our findings significantly enhance the known repertoire of venom gland polypeptides and provide molecular and biochemical evidence for the compartmentalization of this organ into distinct functional entities. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M113.031351, 938–953, 2014.

Animals utilize venoms for many reasons including killing, digestion of prey, protection against predators, and averting competitors. Venom biosynthesis and delivery is achieved through a vast range of structures and mechanisms but commonly includes a venom gland for venom synthesis, processing and storage and a specialized envenomation apparatus that varies strongly with prey preference. Venom glands have diverse evolutionary origins and are often heterogeneous in nature with appearances ranging from kidney-shaped in platypus to sac-shaped in sea urchins and duct-like in bees and cone snails.

The venom gland of predatory marine cone snails (also called venom duct) can measure three to four times the length of the snail's body (1, and this study). At any given time hundreds if not thousands of small peptide conotoxins are biosynthesized in the epithelial cells of the venom gland (2, 3) and secreted into its lumen. Morphological studies indicate that venom is released from the epithelial cells via rupture of cell membranes (4). Peristaltic movement of glandular muscle cells and contraction of the muscular bulb, an organ located at the internal end of the gland, is believed to push the venom toward the pharynx where it is loaded into a harpoon-like radula tooth for injection into the prey (5). Epithelial cell composition, ultrastructure and granule content vary between the proximal and distal portion of the gland with most prominent morphological changes closest to the pharynx (4). Differences

From the [‡]Department of Biology and [§]Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA; [¶]Oral Health Cooperative Research Centre, Melbourne Dental School, and Bio21 Institute, The University of Melbourne, Melbourne, Victoria 3010, Australia; ^{||}Faculty of Veterinary Science, The University of Melbourne, Victoria 3010, Australia; ^{**}The Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia

Received May 28, 2013, and in revised form, January 16, 2014

Published, MCP Papers in Press, January 29, 2014, DOI 10.1074/mcp.M113.031351

Author contributions: H.S., D.G.G., and A.W.P. designed research; H.S. performed research; P.D.V., N.D.Y., E.C.R., and B.M.O. contributed new reagents or analytic tools; H.S., H.H., P.K.B., P.D.V., N.D.Y., and M.Y. analyzed data; H.S. wrote the paper.

EXPERIMENTAL PROCEDURES

in cell morphology were suggested to reflect specializations in venom synthesis, processing, packaging, and secretion (4). Recent transcriptomic and proteomic profiling have begun to unveil regional differences in conotoxin expression and abundance along the gland of *Conus geographus* and *Conus textile* (6–8). In *C. geographus*, a fish-hunting species responsible for at least 30 human fatalities, the majority of conotoxin transcripts are differentially expressed along the venom gland with the most diverse sets of toxins found close to the injection apparatus (7). Mass spectrometric analyses of venom isolated from the mollusk-hunting snail *C. textile* also revealed regional differences in toxin abundance and posttranslational processing (6, 8). However, whether regional expression profiles of conotoxins were reflected by differences in the overall transcriptome and proteome of the venom gland was not addressed.

To comprehensively investigate global expression and abundances of proteins along the length of this complex organ the present study employed transcriptome sequencing and quantitative DIGE analysis combined with mass spectrometric protein identification on four sections of the *C. geographus* venom gland. This systematic approach allowed for the proteomic identification of 408 protein spots corresponding to 342 unique proteins and revealed distinct regional protein abundances across the venom glands of the 3 specimens examined. The visualizing power of DIGE showed that the most abundant proteins are present in multiple isoforms with dramatic changes in regional abundances of individual isoforms. These proteins include two highly abundant proteases of the astacin metalloprotease family, several pore-forming proteins with homology to echotoxins (conoporins), phospholipases of the A_2 family (conodipines) and a number of novel polypeptides of yet unknown function. Enzymes known to play a role in conotoxin folding also exhibited changes in abundances across the gland, including members of the protein disulfide isomerase (PDI)¹ family and prolyl-4 hydroxylase (P4H). Although there was a certain degree of variation between the three specimens the majority of proteins with a significant change between the four sections showed similar patterns in protein abundance across individuals.

This study provides transcriptomic and proteomic evidence for the functional compartmentalization of the venom gland of cone snails significantly complementing and extending earlier work on the morphology and venom content of this unusual organ. It further expands and highlights the genetic diversity of venom gland components enhancing our understanding of cone snail toxin biosynthesis and envenomation.

¹ The abbreviations used are: PDI, protein disulfide isomerase; ASTL, Astacin-like; BVA, biological variance analysis; Cikot, con-ikot-ikot; Cdpi, conodipine; Cporin, conoporin; EDA, extended data analysis; PCA, principal component analysis; P4H, prolyl-4 hydroxylase; UCRP, unknown cysteine rich protein; UP, unknown protein.

Specimen Collection and Tissue Preparation—Specimens of *C. geographus* were collected in Cebu Province, the Philippines. Venom glands from 3 specimens were dissected and divided into four equal-length segments. Shells were between 10–13 cm in length with glands measuring 20, 27, and 30 cm (cut into 5, 6.75, and 7.5 cm segments, respectively).

Venom gland sections were numbered 1–4, starting with the innermost segment connected to the muscular venom bulb. Segment 4 represents the most distal part that connects to the foregut.

Protein Extraction—Frozen venom gland segments were ground into a fine powder at 30 Hz for 2 min using a cryogenic ball mill (MM400, Retsch). The powder was reconstituted in lysis buffer (30 mM Tris, 7 M Urea, 2 M Thiourea, 4% (w/v) CHAPS, pH 8.5) and incubated on ice for 30 min. Proteins were precipitated using the 2D Clean-up Kit following the manufacturer's instructions (GE Healthcare). Dried protein pellets were reconstituted in lysis buffer containing 2% Amidosulfobetaine-14 (Sigma-Aldrich). Proteins were quantified using Bradford reagent (Sigma-Aldrich).

2-Dimensional Fluorescence Difference Gel Electrophoresis (DIGE)—Fluorescent Protein Labeling—Fluorescent labeling and analysis of labeled proteins were carried out under minimal light exposure. Seventy μ g of protein sample were labeled with 300 pmoles of Cy3 or Cy5 fluorescent dye (CyDye™, GE Healthcare) reconstituted in dimethylformamide (see supplemental Table S1 for labeling strategy). Seventy μ g of the internal standard that contained an equal concentration of protein from every sample were labeled with 300 pmoles of Cy2 dye. Labeled samples were incubated for 30 min on ice. Labeling was terminated by adding L-Lysine (Sigma-Aldrich) to a concentration of 1 mM followed by incubation for 10 min on ice.

First Dimension Isoelectric Focusing (IEF)—Isoelectric focusing strips (Immobiline DryStrips, non linear, pH 4–7, 11 cm, GE Healthcare) were rehydrated overnight in Destreak™ Rehydration Solution (GE Healthcare) containing 1% immobilized pH gradient buffer (GE Healthcare). Labeled proteins were pooled as shown in supplemental Table S1 and mixed with an equal volume of 2x sample buffer containing 7 M Urea, 2 M Thiourea, 4% CHAPS and 20 mM dithiothreitol (DTT). Samples were cup loaded and run on the Ettan IPGphor II IEF System (GE Healthcare). Running conditions were 500 V for 1 h at 0.5 kVh, 1000 V for 1 h at 0.8 kVh, 6000 V for 2 h at 7.0 kVh, and 6000 V for 40 min at 0.7–3.7 kVh. Following IEF, strips were reduced in equilibration buffer (75 mM Tris-HCl, 6 M Urea, 30% Glycerol, 2% SDS, 0.002% Bromphenol Blue) containing 65 mM DTT for 15 min followed by alkylation for 15 min in equilibration buffer containing 80 mM iodoacetamide. Second dimension gel electrophoresis was performed on 8–16% Tris-HCl polyacrylamide gradient gels (Criterion, Bio-Rad) for 50 min at 200 V. The differentially labeled co-resolved proteome maps within each DIGE gel were imaged at 100 μ m resolution on a Typhoon 9400 Variable Mode Imager (GE Healthcare) using dye specific excitation and emission wavelengths. Sixteen-bit tagged image files were created in ImageQuant (TL 7.0, GE Healthcare) and exported into DeCyder v7.2 software (GE Healthcare) for statistical analysis using the biological variance analysis (BVA) module. Proteins with statistically relevant changes in abundance (*t* test; *p* < 0.05) were selected for further analysis. Spot matching was manually edited and/or confirmed for all proteins of interest. Principal component analysis (PCA) and Kmeans clustering on these proteins was performed in the extended data analysis (EDA) module using default settings. Relative changes in protein abundance between the 4 sections of the gland are either expressed as fold changes (average ratios) or by comparing normalized gel spot volumes.

Mass Spectrometric Protein Identifications—Gels used for protein spot analyses were prepared as described above with the exceptions that a total of 500 μ g of protein was loaded onto each gel and that

proteins were not fluorescently labeled. For protein visualization, gels were stained with Coomassie Brilliant Blue G-250 (Bio-Rad).

Automated Spot Picking and MALDI-TOF/TOF Analysis—Two preparative gels were selected for MALDI-TOF/TOF analysis. Spot detection was performed using Proteomweaver software (Bio-Rad, Hercules, CA). Individual protein spots were excised from 2D gels using a Proteiner SP spot picker (Bruker Daltonics, Billerica, MA). Cut gel spots were transferred to 96-well plates. In gel tryptic digestion was performed as previously described (9). Three μl of the digested sample was carefully spotted on an AnchorChip containing prespotted matrix (4-hydroxy-cinnamic acid, Bruker Daltonics). After 10 min adsorption, AnchorChips were briefly washed with 0.1% TFA and allowed to dry for 10 min.

Automated mass spectrometric analysis was carried out on an Ultraflex III MALDI-TOF/TOF instrument (Bruker Daltonics). MS was performed using a 25 kV positive reflectron method. Automated MS calibration was carried out every four spots. Peak detection was performed using FlexAnalysis v3.0 (Bruker Daltonics). Tandem mass spectrometry spectra were acquired on parent ions selected by ProteinScape (Bruker Daltonics). Three sets of 100 spectra were accumulated for each parent ion. Peak lists were automatically sent to ProteinScape for removal of calibrants and contaminants including polymers and peptides derived from trypsin and keratin. Peak lists were submitted to Mascot 2.2 (Matrix Science) for peptide mass fingerprint (PMF) and MS/MS ion searches against an in-house database that contained protein sequences generated from the transcriptomes of *Conus bullatus* (10) and *C. geographus* (7) ($n = 3805564$) with the following settings: proteolytic cleavage by trypsin allowing 1 missed cleavage, carbamidomethylcysteine as a fixed modification, methionine oxidation as a variable modification. PMF searches were carried out with a mass tolerance of 100 ppm. Following PMF analysis, automated MS/MS acquisition was triggered for up to six parent ions for proteins that could not be identified by PMF alone, given that the peaks had a “goodness for MS/MS” value greater than 0. PMF results were verified by performing MS/MS on up to three identified peaks. MS/MS ion searches were conducted with a peptide mass tolerance of 100 ppm and a fragment tolerance of 0.8 Da. MS/MS search results were considered genuine if the MS/MS score was greater than 30 or MS score greater than 70. Single peptide identifications were only accepted when the MS/MS score was above the Mascot identity threshold (>38). MALDI-TOF data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (11) with the data set identifier PXD000581 and DOI 10.6019/PXD000581 under accession numbers 33325–33678.

Additional ESI-TripleTOF Analysis—Additional mass spectrometric analysis was performed for a total of 55 proteins that exhibited obvious differences in abundance as judged by 2DGE analysis (see Fig. 1). Spots were manually excised, destained, dehydrated, and trypsin digested as described above. Tryptic peptides were loaded onto a microfluidic trap column packed with ChromXP C18-CL 3 μm particles (300 Å nominal pore size; equilibrated in 0.1% formic acid/5% ACN) at 5 $\mu\text{l}/\text{min}$ using an Eksigent NanoUltra cHiPLC system. An analytical microfluidic column (15 cm \times 75 μm ChromXP C18-CL 3) was then switched in line and peptides separated using linear gradient elution of 0–80% ACN 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 up to 30 MS/MS spectra per second. MS/MS data were searched against the in-house cone snail database using Protein Pilot software (version 3.0, AB SCIEX) with the following selections: iodoacetamide, trypsin gel based identification, biological modifications, thorough ID. The false discovery rate cutoff was set to 5%. Proteins

with > 2 peptides with an individual peptide score of ≥ 99 were regarded as genuine identifications.

Interrogation of the Regional Venom Gland Transcriptome—The recently published transcriptome of *C. geographus* was interrogated to determine regional differences in gene expression patterns along the venom gland ((7), Data is available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>) under accession numbers SRR503413, SRR503414, SRR503415 and SRR503416). Briefly, as outlined in the original study (7), the transcriptomes of 4 equal-length venom gland segments (pooled from four specimens) *C. geographus* were independently sequenced on the Roche Genome Sequencer FLX Titanium platform. A total of 167,211, 238,682, 186,398, and 199,680 high-quality reads were generated for the Proximal (segment 1 or P), Proximalcentral (segment 2 or PC), Distal-central (segment 3 or DC) and Distal segments (segment 4 or D), respectively. The average read length was 425.8 bp with an N50 read length of 580 bp. Reads were pooled and assembled using Mira3 software (12) to generate a reference transcriptome database containing 49,515 contigs of 20.8 Mbp in length after removal of redundancies (N50: 576 bp). Raw reads generated from each segment were aligned to the reference database using the Burrows-Wheeler Alignment tool (13), resulting in 98.7%, 99.3%, 99.1%, and 99.2% of aligned reads for the four sections. Annotations were performed using BLASTX (14) and InterProScan (15).

Normalization of gene expression levels was performed according to the method developed by Robinson and Oshlack (16). For each contig, a p value was calculated using a chi-square test under the null hypothesis of equal expression across the gland. Expression analysis was only performed for transcripts with ≥ 10 reads and $p < 0.01$.

RESULTS

Distinct Abundances of Proteins in the Venom Gland—Proteins were prepared from four sections of the venom glands of 3 specimens of *C. geographus*, separated by 2DGE, stained with Coomassie Brilliant Blue or fluorescently labeled for DIGE analysis. Major differences in protein abundance along the gland were apparent with most prominent changes between most distant regions (Fig. 1). Fewest changes were observed between section 1 and 2 indicating that these regions are functionally similar entities. With a few exceptions, protein abundances were strikingly similar between the three individuals with slight differences in the abundance of some protein spots and in the onset of change. Differences in specimen size, age, and the length of venom glands (ranging from 20–30 cm) are likely to have contributed to these slight variations.

BVA analysis revealed that a total of 157 protein spots exhibited significant differences in protein abundance across the four sections with least changes between section 1 and 2. Indeed, PCA analysis of spot maps could not satisfactorily resolve these two sections because of high spot pattern similarities whereas section 3 and 4 were classified into distinct groups (Fig. 2).

Gel analysis revealed 8 groups of proteins with obvious changes in abundances across the gland (Fig. 1, boxed red). In order to identify these proteins a total of 55 protein spots were excised from preparative gels (B2 - B4) and subjected to in-gel tryptic digestion and mass spectrometric analysis on a

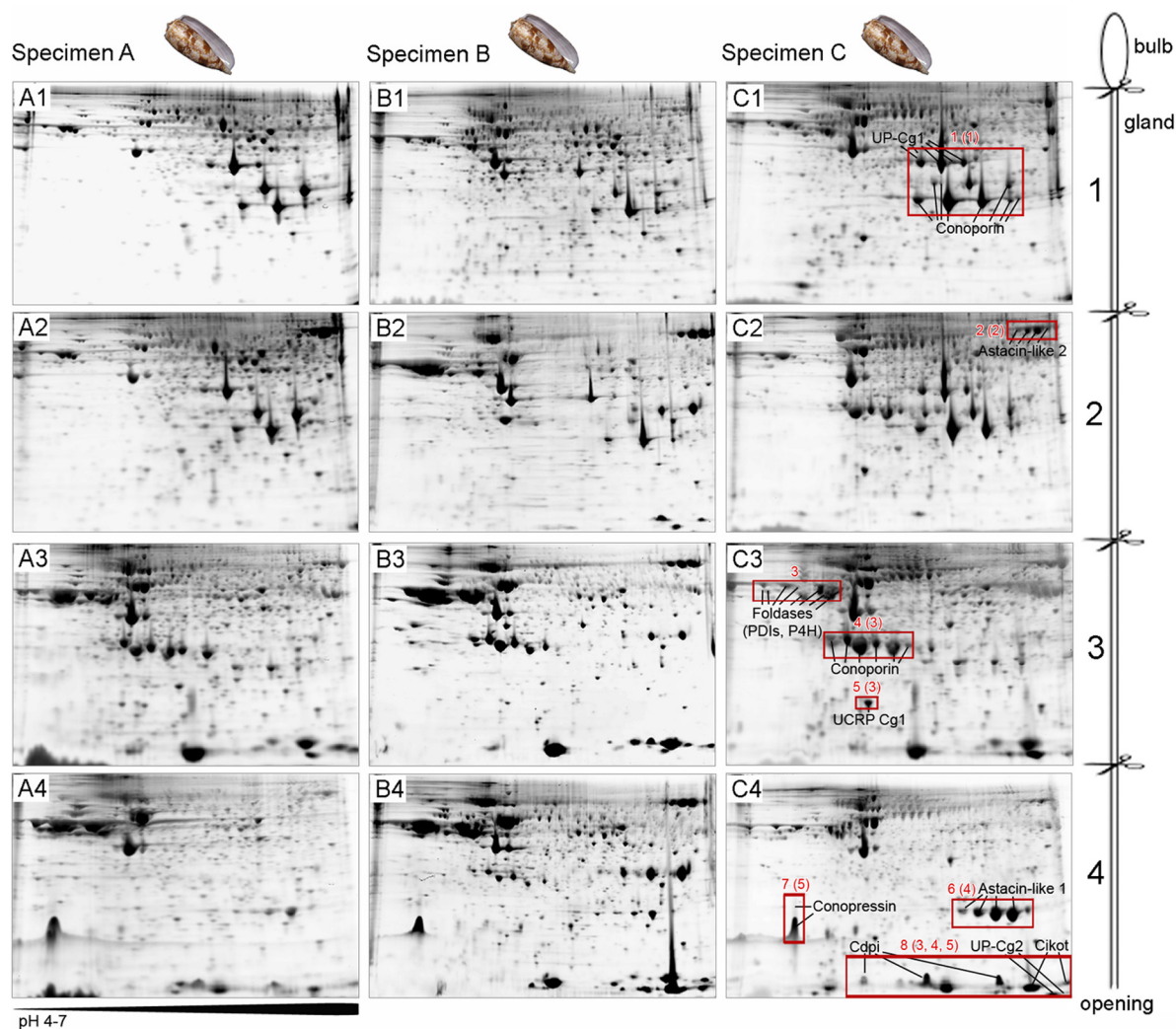


FIG. 1. 2D gel images of proteins extracted from 4 venom gland sections (1–4) of three specimens of *Conus geographus* (A, B and C). Proteins were stained with Coomassie for spot visualization. Gel spots were excised, digested with trypsin and analyzed by LC-MS/MS and MALDI TOF-MS/MS. Mass spectrometric data were searched against an in-house *Conus* database for protein identifications using Protein Pilot and Mascot software. *Kmeans cluster analysis* was performed on fluorescently labeled gels in the extended data analysis module of DeCyder. Protein groups of interest are boxed and group IDs are provided followed by cluster IDs in parentheses. Details on protein identifications are provided in Table I and supplemental File S1. The following proteins were identified (corresponding gel spot numbers are shown in parentheses): UP-Cg1 (3b, 22b, 34b, 87c, 173c, 174c, 175c, 180c, 181c), Conoporphin (group 1: 12a, 4b, 7b, 8b, 11b, 20b, 33b, 41b, 56b, 77b, 78b), Astacin-like 2 (6a, 8a, 9a, 16b, 18b, 55b, 58b), Foldases (1a, 2a, 3a, 4a, 36a, 38a, 43b, 93b, 290c), Conoporphin (group 4: 24a, 25a, 26a, 27a, 28a, 29a, 19b, 25b, 65b), UCRP-Cg1 (401c), Astacin-like 1 (50a, 51a, 52a, 397d, 408d, 409d), Conopressin (35d, 388d), Con-ikot-ikot (35a, 55a), Conodipine (53a, 390d, 396d, 401d), UP-Cg2 (418d, 419d).

Triple TOF LC-ESI-MS/MS instrument. An additional set of 355 protein spots excised from gel A2 and A4 were identified by MALDI TOF-TOF mass spectrometry. All protein identifications are provided in supplemental File S1.

Mass spectrometric analysis led to the identification of all gel spots of interest within the 8 groups of differentially expressed proteins (Fig. 1, Table I and supplemental File S1). The majority of these proteins are likely to function in venom synthesis and maturation or form part of the injected venom mixture. With the exception of protein group 3, BVA analysis confirmed differential abundances of these proteins at a p value of < 0.05 and fold changes of between -23 and 481

(Fig. 3). Based on their regional abundances protein groups were classified into distinct clusters. *Kmeans cluster analysis* was performed in the EDA module of DeCyder with q values provided in Fig. 3.

Group 1 comprised different isoforms of conoporphin and a novel protein of yet unknown function. Proteins identified in this group were most abundant in the first two sections and classified into cluster 1 (Fig. 1 and 3Ci). Conoporphins are proteins with homology to echotoxins isolated from the salivary gland of the marine gastropod *Monoplex echo* and actinoporphins from the nematocysts of sea anemones. Here, gel analysis revealed the presence of several isoforms of conoporphin

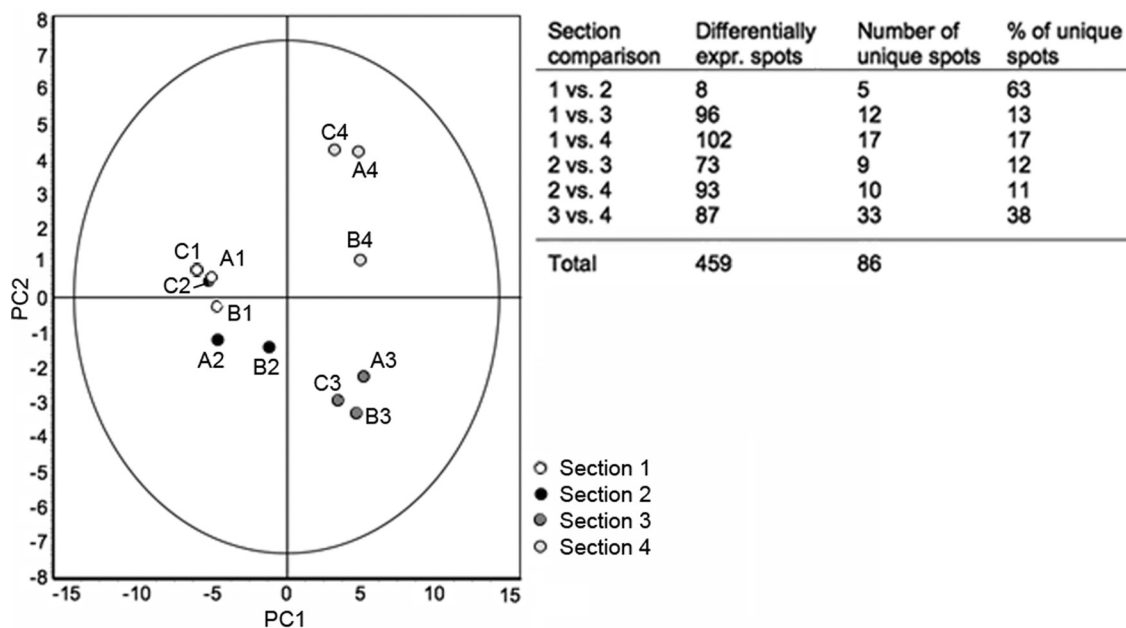


FIG. 2. Principal component analysis (PCA) showing distinct grouping of section 3 and 4 and an overlap for section 1 and 2. Section C2 and B1 were misclassified into section 1 and 2, respectively indicative of functional similarities between these adjacent sections. PCA was performed on all differentially expressed proteins ($n = 157$, t test: $p < 0.05$) in the extended data analysis module of DeCyder using default settings. Numbers of proteins that showed significant changes in abundance as determined by biological variance analysis between the four sections are shown in the table.

with distinct migration patterns. Little variation was observed between the three individuals tested rendering these spots likely to be real isoforms rather than technical artifacts. Transcriptome mining indeed identified 9 transcripts encoding full-length proteins with differences in size, amino acid composition and isoelectric point (Cporin-Cg1 - 9, Fig. 4). All translated sequences contain an N-terminal signal peptide and the cytolysin/lectin domain characteristic for actinoporin-like proteins (Fig. 4, arrows).

Cluster 1 also contained a protein of yet unknown function. This protein was tentatively named 'Unknown Protein *Conus geographus* 1' (UP-Cg1). Gel analysis identified several isoforms of this protein 2 of which showed a significant decrease in abundance toward the last gland section (Fig. 3 Ci). However, only 1 transcript could be identified in the transcriptome database indicating that posttranslational modifications may have accounted for the presence of 2 isoforms or that sequence data is incomplete. Sequence analysis using several different algorithms did not identify any domain other than an N-terminal signal peptide (Fig. 5, top panel). Mature UP-Cg1 contains four cysteine residues and is 339 amino acids long (362 containing the signal peptide) with a predicted molecular mass of 37936 Da.

Group 2 proteins were identified as different isoforms of a zinc metalloprotease of the astacin family and were classified into cluster 2 with highest abundance in sections 2 and 3 (Fig. 1 and 3Cii). Because of high interspecimen variation clustering was comparatively weak (q value: 42). Interestingly, a

homologous family of proteins was identified in a different protein group/cluster (group 6, cluster 4, see below).

Group 3 showed obvious differences in protein abundance with high variations between the three specimens tested. This group was most abundant in section 3 and 4 of specimen A and B but not C. Although protein spots belonging to this group could be resolved on fluorescently labeled gels, gels used for mass spectrometric identification did not yield sufficient resolution power for unambiguous protein identifications. However, it can be noted that most protein spots identified in this gel area belonged to proteins known to play a role in conotoxin folding and modification (17, 18), that is, members of the PDI family, several heat shock proteins and P4H (Fig. 1, supplemental File S1). Additionally, statistical analysis of these protein spots was hampered by high interspecimen variation. Consequently, proteins could not be grouped into clusters and statistical analysis is not provided.

Similar to group 1, **group 4** proteins were identified as different isoforms of conoporin. In contrast, this group migrated at a lower pI and was classified into cluster 3 with highest abundance in section 3 followed by a slight decrease in section 4 (Fig. 1 and 3Ciii). Proteomic peptide matching assigned most group 1 tryptic peptides to gene transcripts Cporin-Cg1 and Cg2 whereas group 4 peptides predominantly matched to Cporin-Cg3 (Table I, Fig. 4 and Supplemental Fig. S2). All 3 transcripts encode proteins with low theoretical pIs (6.34–6.65), however Cporin-Cg3 gel spots migrate lower than its predicted pI indicative of the presence

TABLE I

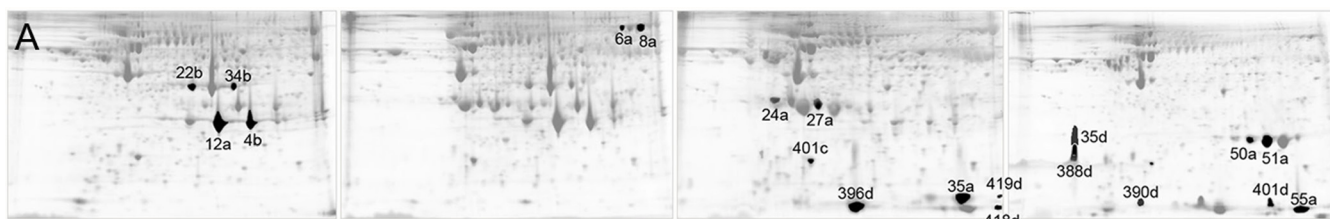
Proteomic identification of differentially abundant proteins in the venom gland of *C. geographus*. The order is according to Figs. 3B and 3C. Sequence coverages are based on the best assembled contiguous sequence (contig) without N-terminal signal sequences. In some instances these contigs do not represent full sequences (marked with *). Where multiple protein IDs are provided (e.g. con-ikot-ikot Cg4 and Cg5 for spot 35a) peptide matches can be aligned to two different isoforms that cannot be distinguished based on proteomic data. Protein Pilot and Mascot scores are provided for ESI-Triple TOF and MALDI-TOF/TOF data, respectively

Protein spot	Protein name	Sequence coverage (%)	No of peptides	Nonredundant peptides	Score, mascot (M); protein pilot (PP)	Molecular mass (Da)	Cluster
22b	Unknown Protein 1	21	6	6	362 (M)	37936	1
34b	Unknown Protein 1	21	6	6	424 (M)	37936	1
4b	Conoposin Cg1	20	3	3	236 (M)	24574	1
	Conoposin Cg3	11	1	1	43 (M)	24627	1
12a	Conoposin Cg1	62	38	12	≥99 (P)	24574	1
	Conoposin Cg2	88	1548	58	≥99 (P)	24743	1
	Conoposin Cg3	37	7	5	≥99 (P)	24744	1
	Conoposin Cg4	10	5	1			
6a	Astacin-like 2	31*	106	20	≥99 (P)	Incomplete	2
8a	Astacin-like 2	58*	370	56	≥99 (P)	Incomplete	2
24a	Conoposin Cg2	26	3	2	≥99 (P)	24722	3
	Conoposin Cg3	76	146	14	≥99 (P)	24627	3
	Conoposin Cg4	9	5	2	≥99 (P)	24949	3
27a	Conoposin Cg3	66	432	18	≥99 (P)	24722	3
401c	Unknown Cysteine-Rich Protein 1	13	2	2	98 (M)	19277	3
35a	Con-ikot-ikot Cg4 and Cg5	31	51	8	≥99 (P)	9468	3
	Con-ikot-ikot Cg6	19	2	1	≥99 (P)	10640	3
419d	Unknown Protein 2	28	2	2	283 (M)	8384	3
418d	Unknown Protein 2	28	2	2	276 (M)	8384	3
396d	Conodipine Cg1 and Cg2	11	1	1	56 (M)	15197	3
401d	Conodipine Cg1 and Cg2	11	1	1	56 (M)	15197	4
390d	Conodipine Cg2	10	1	1	51 (M)	15197	4
50a	Astacin-like 1	54*	102	24	≥99 (P)	32066	4
51a	Astacin-like 1	64*	521	62	≥99 (P)	32066	4
35d	Conopressin	26*	1	1	47 (M)	Incomplete	4
388d	Conopressin	40*	1	1	56 (M)	Incomplete	5
55a	Con-ikot-ikot Cg6 and Cg7	49	77	16	≥99 (P)	10519	5

of one or more acidic modifications. The identification of conoposins with acidic pIs is surprising as actinoposin-like proteins are generally very basic with pIs above 9 (19). Transcriptome analysis identified several sequences encoding basic conoposins (Cporin-Cg4 - 9, Fig. 4), however, these proteins could not be resolved on pH 4–7 strips used here.

Group 5 only comprised 1 gel spot that was identified as a novel yet unknown cysteine-rich protein and classified into cluster 3 (Fig. 1 and 3Ciii). This protein was tentatively named ‘Unknown Cysteine-Rich Protein *Conus geographus* 1’ (UCRP-Cg1). A transcript encoding the full-length open reading frame of this protein was retrieved from the transcriptome database. A homologous transcript was also identified in the venom gland transcriptome of *C. bullatus* and is shown for comparison (Fig. 5, bottom panel). Sequence analysis using several different algorithms did not identify any domain other than an N-terminal signal sequence. Mature UCRP-Cg1 contains 12 cysteine residues (14 including the signal peptide) and is 171 amino acids long with a predicted molecular mass of 19277 Da for the linear protein (Fig. 5, bottom panel).

Proteins belonging to **group 6** were identified as different isoforms of a zinc metalloprotease of the astacin family. The same protein family was also identified for group 2, however, the two protein groups show distinct gel migration patterns and were classified into different clusters. Unlike group 2 that showed highest abundance in section 2, group 6 astacin-like proteases belong to cluster 4 and are mostly found in section 4 with very low abundance in the first two sections (Fig. 1 and 3Civ). Interestingly, tryptic peptides from group 2 matched to a full-length transcript from the *C. geographus* venom gland whereas peptides obtained for group 6 corresponded to a partial sequence retrieved from the *C. bullatus* transcriptome. This allowed for the assignment of these two groups of astacin-like proteases to distinct gene transcripts. The two partial transcripts were tentatively named ‘Astacin-like 1 isoform 1 from *Conus geographus*’ (ASTL1-Cg1) and ‘Astacin-like 2 isoform 1 from *Conus bullatus*’ (ASTL2-Cb1) (Fig. 6). Although the two proteins are clearly homologous (maximum identity: 39%, E value: 9e-42) they significantly differ in their amino acid composition and domain organization. Both proteins



B

Protein (Gel Spot)	Section 2	Section 3	Section 4	Cluster
Unknown P1 (22b)	-1.1	-21.0	-23.2	1
Unknown P1 (34b)	-1.1	-3.5	-7.9	
Conopirin (4b)	-1.1	-9.2	-14.7	
Conopirin (12a)	-1.0	-8.5	-19.7	
Astacin-like 2 (6a)	3.8	1.9	1.2	2
Astacin-like 2 (8a)	3.1	1.3	-1.5	
Conopirin (24a)	2.4	3.8	-1.3	3
Conopirin (27a)	10.8	22.7	3.6	
Unknown CRP1 (401c)	1.1	7.3	2.0	
Con-ikot-ikot (35a)	1.9	150.3	44.8	
Unknown P2 (419d)	6.2	34.8	12.9	4
Unknown P2 (418d)	7.3	65.5	46.4	
Conodipine (396d)	3.8	480.7	126.7	5
Conodipine (401d)	-1.0	3.5	9.0	
Conodipine (390d)	-1.4	-1.5	1.9	
Astacin-like 1 (50a)	-1.2	-4.9	8.5	
Astacin-like 1 (51a)	-1.4	-1.6	127.9	
Conopressin (35d)	-1.2	1.3	14.4	
Conopressin (388d)	1.4	1.3	33.4	
Con-ikot-ikot (55a)	1.7	3.8	14.6	

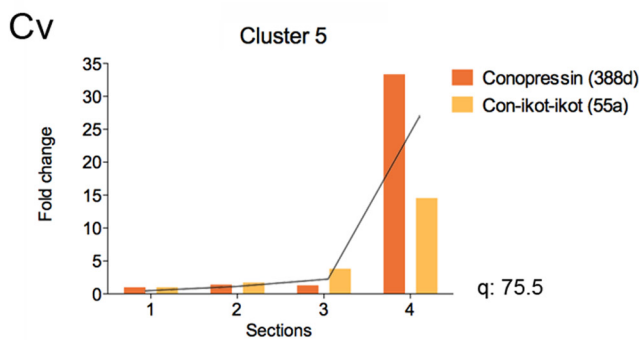
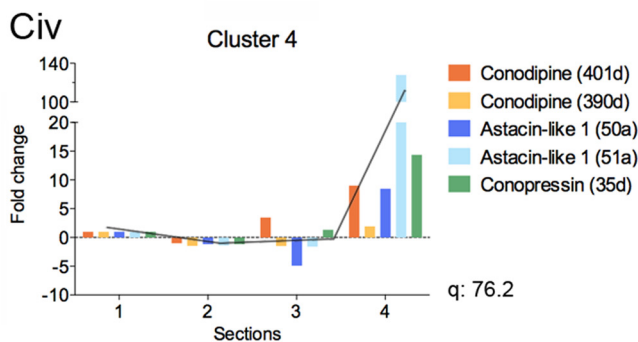
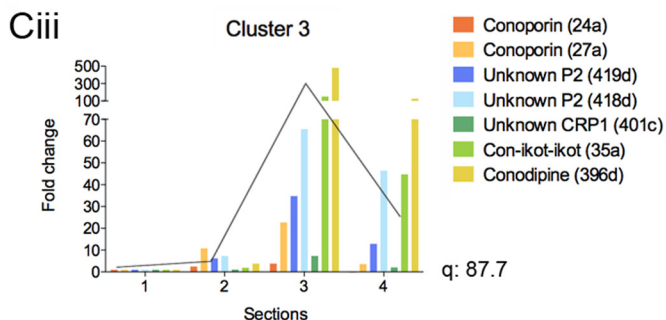
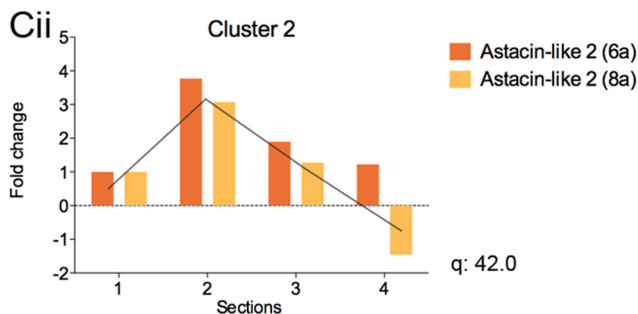
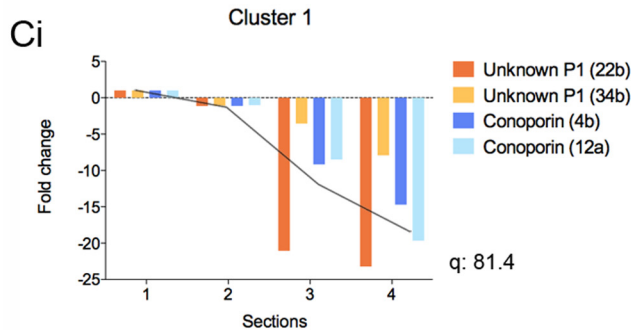


FIG. 3. Differential expression analysis of selected proteins and protein isoforms as determined by biological variance analysis (BVA). Corresponding 2D gel spot IDs are shown in A (label a, b, c, and d correspond to gel A2 part 1, A2 part 2, A4 and B2–4, respectively (see supplemental File S1 for details)). Values of average fold changes and classification into clusters are provided in B. Values for section 1 were set to 1. Kmeans cluster analysis was performed in the extended data analysis module of DeCyder using default settings. Panels Ci - Cv show regional fold changes in protein abundance across the four sections sorted after clusters. q values of cluster analyses are provided next to graphs. Lines show averages of all proteins spots grouped in each cluster. All proteins except for those in cluster 2 showed significant regional abundances ($p < 0.05$, Student's *t* test).

contain the M12A peptidase domain of the astacin family (IPR001506) with the zinc-binding motif HEXXH (Fig. 6, boxed). Interestingly, ASTL2-Cb1 also comprises three Shk toxin domains (IPR003582) C-terminally of its metalloprotease

domain. These domains share sequence homology with ShK, a cysteine-rich toxin isolated from the sea anemone *Stichodactyla helianthus* (20). ShK is a potent inhibitor of K^+ channels. ShK-like domains have been identified in several other asta-

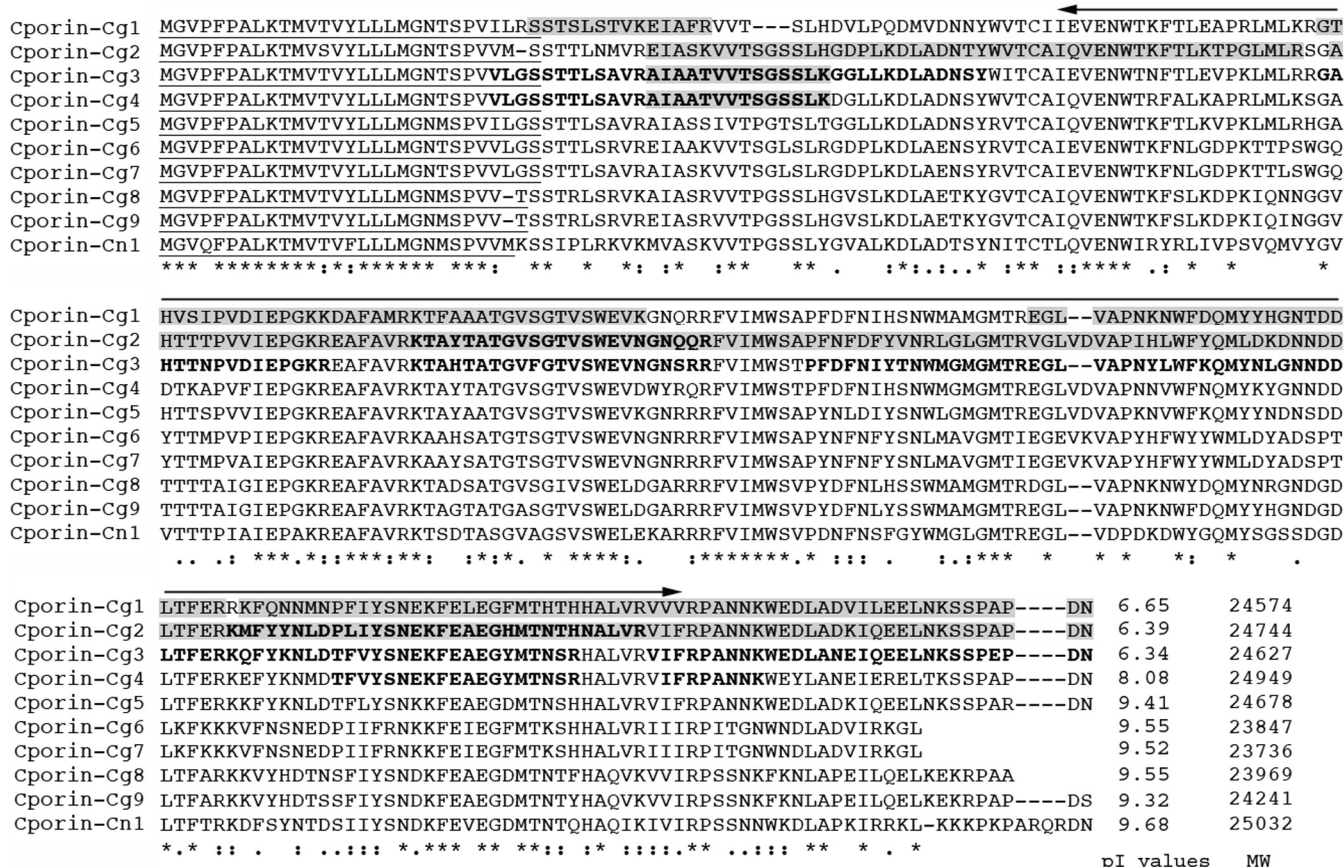


FIG. 4. Comparative alignment of conoporphin (Cporin) sequences identified in the venom gland of *C. geographus* (Cporin-Cg1 - 9; GenBank accession number: GAJN01000001-GAJN01000009). Conoporphin-Cn1 from *C. consors* is shown for comparison (P0DKQ8). Multiple sequence alignment was performed using MAFFT auto alignment (version 7) (42). Signal sequences and domains were predicted using InterProScan (43). Signal sequences are underlined. Arrows depict position of the cytolysin/lectin domain (IPRO15926). Tryptic peptides identified for cluster 1 and 3 are highlighted in gray and shown in bold, respectively. Peptides for individual gel spots identified as conoporphins are provided in supplemental Fig. S2. Isoelectric points (pI) and molecular weights (MW) of the mature proteins (without signal sequences) are shown next to sequences. Amino acid conservations are denoted by an asterisk (*). Full stops (.) and colons (:) represent a low and high degree of similarity, respectively.

cin-like proteases including human matrix metalloprotease 23, a functional protease and K⁺ channel modulator (21).

Gel spots belonging to **group 7** were identified as conopressin prohormones and classified into cluster 4 and 5 with highest abundance in the last segment. Protein spot volumes either continuously increase (Fig. 3Cv) or increase after a slight drop in section 2 and/or 3 (Fig. 3Civ). Preparative gel analysis suggested vertical streaking of this spot (Fig. 1, group 7), however, the visualization power of DIGE revealed the presence of two distinct protein spots, both identified as conopressin prohormones by mass spectrometric analysis. Conopressins are short peptide hormones that share sequence homology with the vasopressin peptide family. Vasopressins are activated via release from a larger cysteine-rich prohormone (neurophysin) which in turn serves as a carrier protein for hormone delivery (22). Conopressins have been identified in the venoms of several cone snail species with no cDNA sequence data available elucidating their precursor orga-

nization (23, 24). Here, venom gland transcriptome mining led to the identification of a partial sequence that suggests the typical prohormone organization of the vasopressin family (Supplemental Fig. S1). Interestingly, this sequence shows homology to conophysin-R, a polypeptide belonging to the neurophysin family that was identified from the venom of *Conus radiatus* (25). Additionally, gel spots identified as conopressin migrated at ~15 kDa suggesting that this peptide is indeed translated as a larger conopressin/conophysin prohormone precursor.

Group 8 comprised proteins migrating at the bottom of the gel with a molecular mass of ~8–12 kDa. This group contained several proteins and protein isoforms that belonged to three different clusters; Conodipines (cluster 3 and 4); conikot-ikots (cluster 3 and 5) and a polypeptide of yet unknown function (cluster 3).

Conodipines are members of the phospholipase A₂ (PLA₂) family that hydrolyze ester bonds in phospholipids. Conodipine-M was the first PLA₂ identified in cone snail venom (*C.*

UP-Cg1	<u>MTSPFAGLALLCV</u> CLSLTVPVSGSRLKLVKLTQEFLONTKAYRLTADFEADSRPEAQWLYGHYDVLDPSPGSILTYGHDPNPMKGS ATLTLKDDAVKGVFLSLGLDLSHFLEWNLGGTDDLNVQLDLPLSTVSPGTDLAYNSGKAIYLTIAVKEHRGETDFGPYFNVYLS GDVIYDRVWFFRYMGRPVFEESESKNGVTTSVKFLTDYRPISGIFSGKLYGILNHNANAPVTRLIEHFTVKVHAASTAGPFPDNAVV ARHEDPTICSRDVSYQCYLGCCTGFGNDLDSVKIYQEDQYGWRKLLPAKVSERVQVAFRRRTVKVKVDTGPLPKSKFVCEATLKGGRV EDRFILSIEN	MW 37936 pI 6.02 Cys 4
UP-Cg2	<u>MGATLVTKLLLAALLLGL</u> CHEMAANPEAWDECYERMHMLIAPTWHAMYNCHHEILEDKRAALVCLKQKQVGLVHLDHGG	
UP-Cb2	<u>MGATLVTKLLLVAAALLLGL</u> CYEMAANPEAWDECYERMHMLIAPTWHAMYNCHHELLEDKRAALVMKQKQVGLVHLDHGG *****.*****:*****:*****:*****:*****:*****:*****	
UP-Cg2	<u>WVQHVKLLKYLGVNN</u>	MW 8384
UP-Cb2	<u>WVQQVKLLKYLGM</u> SH ***:*****:..	pI 7.61 Cys 2
UCRP-Cg1	<u>MKTL</u> CGNAVCGLLTVMLISMGKANGQNGWFRKFDVQGLNNAVDFNQIFEPSYTOHLCSQVQPLEGCIGAVRSECPDL	
UCRP-Cb1	<u>MKVL</u> CGTAVCGLLTVLLISMGKANGQVCGWYRKFQVQGLRNLRVDFNQIFVPSYTOQLCSLLQPLEGCIGAVRSQCHDL *.***.*****:***** ***:*****. *: ***** ***:*** :*****:*** **	
UCRP-Cg1	<u>AVFQGGKGLAIDTVL</u> ASLSRI ^C VGNGGDVVPASGGRCPGTESKFRRLGNCHAPMQMNLQESQMCQPMDEIITCMEGRLL	
UCRP-Cb1	<u>TVFQGGKGLAINTV</u> VASLSQ ^I CAGNGGERVTPASGGNCSGEESEFRQLGNCHPEMQMNLRESQMCQPTSEVIACMEGQL :*****:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*.***:*	MW 19277 pI 7.05 Cys 12
UCRP-Cg1	<u>QFRP</u> CSWMTTFLADRQVFWIDFERALWVRYECAGR	
UCRP-Cb1	<u>QTRP</u> CSWMTAFLADRQVFWIDFERALWVQYECAGR * *****:*****:*****:*****:*****	

Fig. 5. Sequences and comparative sequence alignments of novel proteins of yet unknown function (GenBank accession number GAJN01000010-GAJN01000012). DIGE analysis identified regional differences in abundances for 3 unknown proteins with no known domains other than an N-terminal signal peptide (underlined). Proteins were tentatively named Unknown Protein 1 and 2 from *Conus geographus* (UP-Cg1 and UP-Cg2, top and middle panel) and Unknown Cysteine-Rich Protein from *Conus geographus* 1 (UCRP-Cg1, bottom panel). Tryptic peptides sequenced by mass spectrometry are boxed. Homologous sequences were identified in *C. bullatus* (Cb) and are shown for comparison. Multiple sequence alignment was performed using MAFFT auto alignment (version 7) (42). Cysteine residues are shown in bold gray. White arrow depicts a putative triple basic cleavage site for UP-Cg2. Isoelectric points (pI), molecular weights (MW) and number of cysteines are provided for mature proteins (without signal sequences). Amino acid conservations are denoted by an asterisk (*). Full stops (.) and colons (:) represent a low and high degree of similarity, respectively.

magus (26)) followed by the recent discovery of a homologue in *C. consors* (conodipine-Cn (27)). Here, conodipine tryptic peptides matched to several gel spots indicative of the presence of multiple isoforms. Transcriptome analysis revealed 10 unique conodipine sequences in *C. geographus* with differences in sequence composition and intercysteine loop spacing (Fig. 7). This provides the first complete sequences for this *Conus* protein and identifies a previously unknown diversity of this gene family. All sequences contain an N-terminal signal peptide. Based on sequence similarities these signal sequences can be assigned into two groups (group 1: Cdpi-Cg 1–5 and group 2: Cdpi-Cg 6–10, Fig. 7). Most variability between the different isoforms is observed in the sequence located between the a and b chain and the intercysteine loop between Cys8 and Cys9 pointing toward a less conserved function of these regions. *Conus* conodipines contain the His/Asp catalytic dyad characteristic for the secreted PLA₂ family. Interestingly, Kmeans cluster analysis classified the three gel spots identified as conodipine into two different clusters (3 and 4) demonstrating isoform specific distribution patterns across the gland (Fig. 3 Ciii and iv). Conodipines share very little homology with other members of the PLA₂ family, including those sequenced from snake, spider and

cephalopod venom. Mature conodipine-M consists of an a and b chain interlinked by one or more disulfide bonds. All sequences retrieved from the *C. geographus* transcriptome except for one (Conodipine-Cg2) suggest similar 3-dimensional structures (Fig. 7). Unlike the nine other conodipine sequences, conodipine-Cg2 contains two instead of three cysteines in the putative b chain and does not contain a strong basic cleavage site between the two chains suggesting that it may form a homodimer similar to several PLA₂ members identified in snake venom (28).

Con-ikot-ikot was first isolated from the venom of the fish-hunter *Conus striatus* (29) with a homologous protein subsequently found in the venom of *Conus purpurascens* (p21a, (30)). The active *C. striatus* polypeptide is a tetramer (a non-covalent dimer of a covalent dimer) that targets the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor, a subtype of the ionotropic glutamate receptor (29). Similar to observations made for conodipines, tryptic con-ikot-ikot peptides derived from several gel spots suggesting that *C. geographus* utilizes a panel of isoforms of this polypeptide for envenomation. Venom gland transcriptome mining indeed revealed the presence of at least six isoforms of this protein with differences in size, amino acid composition and

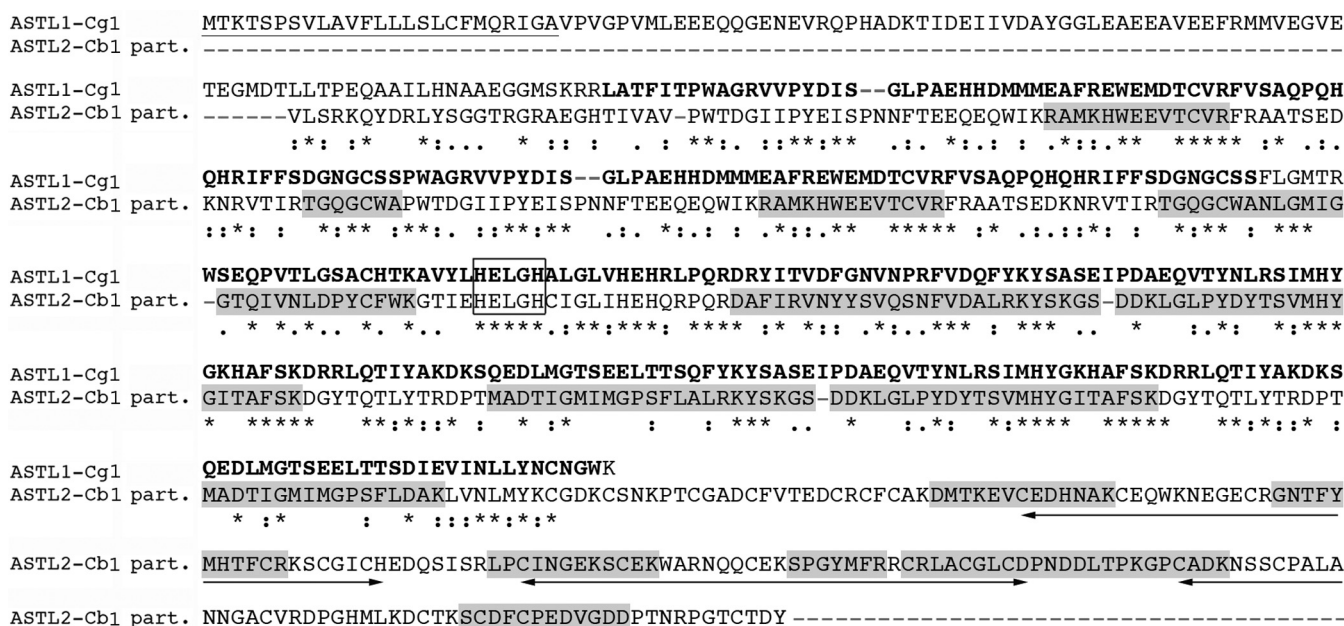


FIG. 6. Comparative alignment of astacin-like metalloproteases identified in the venom gland of *C. geographus* (GenBank accession number GAJN01000013). Tryptic peptides sequenced from group 2 and 6 (see Fig. 1) matched to 2 different sequences retrieved from the venom gland transcriptome of *C. bullatus* and *C. geographus*, respectively. Peptide sequences from group 2 and 6 are highlighted in gray and shown in bold, respectively. Based on sequence similarities to other astacin-like metalloproteases, proteins were named ‘Astacin-like 1 isoform 1 from *Conus geographus*’ (ASTL1-Cg1, full-length) and ‘Astacin-like 2 isoform 1 from *Conus bullatus*’ (ASTL2-Cb1, partial). The 2 proteins share 39% identity and both contain the M12A peptidase domain of the astacin family with the zinc binding motif HEXXH (boxed). ASTL2-Cb1 also comprises 3 Shk toxin domains indicated by black arrows. Sequence alignment was performed using MAFFT auto alignment (version 7) (42). The N-terminal signal sequence of ASTL1-Cg1 is underlined. Amino acid conservations are denoted by an asterisk (*). Full stops (.) and colons (:) represent a low and high degree of similarity, respectively.

cysteine framework (Fig. 8) uncovering a yet unknown genetic diversity of this venom component. Like conodipines, Kmeans analysis revealed two distinct clusters for con-ikot-ikot isoforms indicative of isoform-specific regional expression (Fig. 3 Ciii and v).

Group 8 also comprised two gel spots that were identified as a novel yet unknown protein tentatively named ‘Unknown Protein *Conus geographus* 2’ (UP-Cg2). A transcript encoding the full-length open reading frame of this protein was retrieved from the transcriptome database (Fig. 5, middle panel). A homologous sequence was also identified in the venom gland transcriptome of *C. bullatus* and is shown for comparison (UP-Cb2). Sequence analysis using several different algorithms did not identify any domain other than an N-terminal signal sequence. However, UP-Cg2 displays a sequence that potentially resembles a novel conotoxin. The signal peptide is followed by a putative toxin sequence with two cysteine residues that could form one disulfide bond. A triple basic putative proteolytic cleavage site is located C-terminally of the predicted toxin sequence. A propeptide sequence is missing.

Correlation of Protein Abundance With Gene Expression—To investigate whether regional changes in the proteome are reflected by differences in mRNA transcription the transcriptomes of the 4 sections of the *C. geographus* venom gland were interrogated. The vast majority of aligned reads

within annotated transcripts comprised conotoxin sequences (88%, (7)). This highlights the specialized function of the venom gland in toxin biosynthesis. Significantly fewer reads were obtained for larger polypeptides with the highest number sequenced for con-ikot-ikot (8551 reads). A total of 148 gene transcripts showed significant differences in expression across the four sections (minimum number of reads: 10, $p < 0.01$, supplemental File S2). The majority were ribosomal proteins (34%) with highest expression in segment 2 indicating high protein turnover rates in this region.

Several transcripts were highly expressed but their encoding proteins could not be identified by proteomic analysis. These include proteins previously identified in animal venoms such as hyaluronidase (*C. consors*, (31)), venom basic protease inhibitor 1 (*Vipera ammodytes*, (32)) and the protease inhibitor bitisilin-3 (*Bitis gabonica*, (33) (supplemental File S2)). Future studies are needed to determine whether these proteins form part of the venom of *C. geographus*.

Because of limitations with unambiguous protein identifications of DIGE gel spots, a global comparison between protein abundance and gene expression was not performed and correlation analysis focused on proteins with highest changes in abundances across the gland (see protein groups 1 - 8 above). In order to investigate the correlation between protein and gene expression for these proteins and gene transcripts,

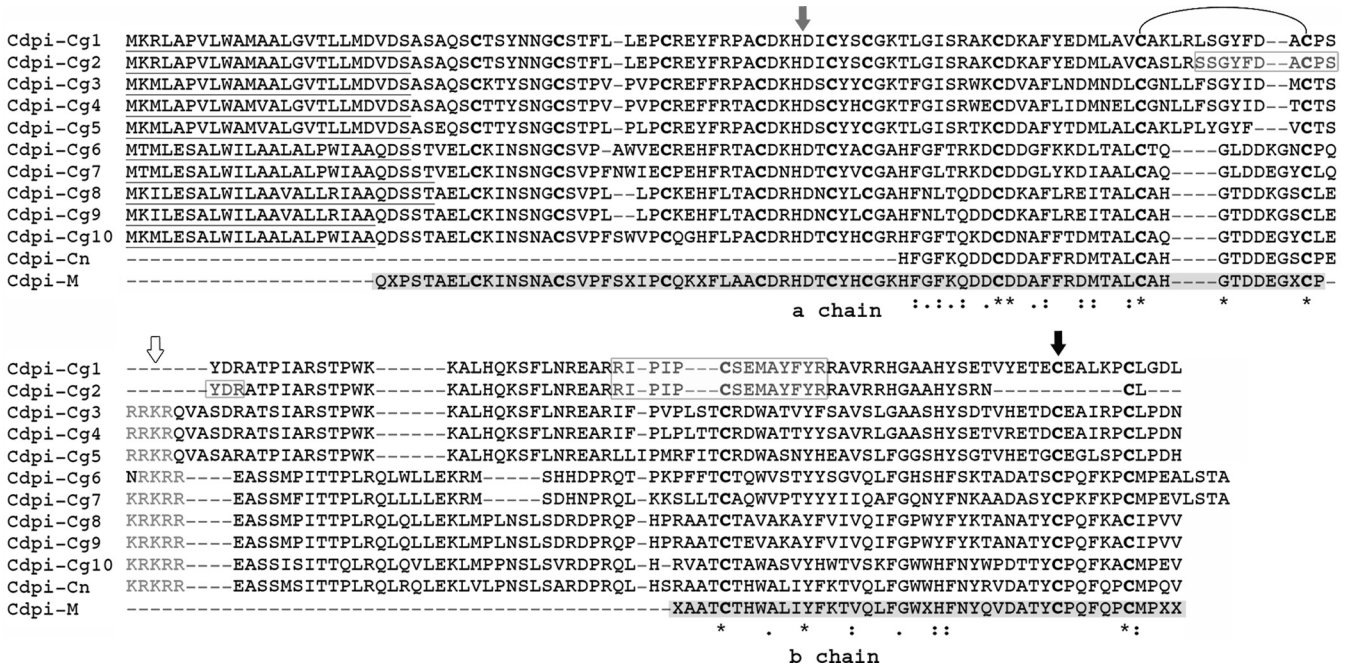


FIG. 7. Comparative alignment of conodipine sequences identified in the venom gland of *C. geographus* (Cdpi-Cg1 - Cg10; GenBank accession number GAJN01000014-GAJN01000023). Multiple sequence alignment was performed using MAFFT auto alignment (version 7) (42). Cdpi-Cn from *C. consors* and Cdpi-M from *C. magus* are shown for comparison (partial sequences). Tryptic peptides identified by mass spectrometry are boxed. The a and b chain identified for Cdpi-M are highlighted in gray. The gray arrow shows the conserved HD motif of the phospholipase A₂ family. The black arrow points at the cysteine residue missing in Cdpi-Cg2. Loop shows the highly variable inter-cysteine loop located between Cys8 and Cys9. Isoelectric points and molecular weights are not provided because proteolytic processing into a covalently linked a and b chain is likely to occur for all sequences but Cdpi-Cg1 and 2 that lack a strong basic cleavage site following the putative a chain (basic sequence shown in light gray). Signal sequences are underlined. Amino acid conservations are denoted by an asterisk (*). Full stops (.) and colons (:) represent a low and high degree of similarity, respectively.

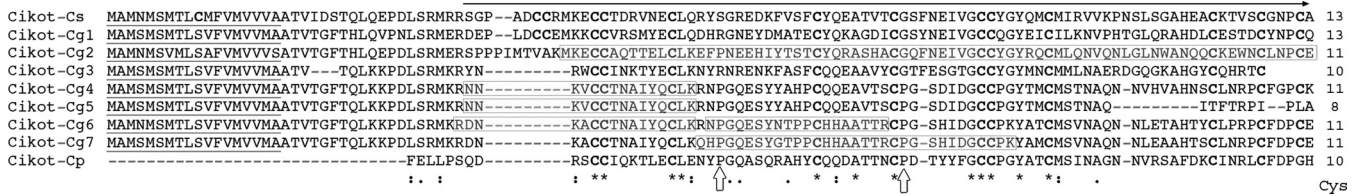


FIG. 8. Comparative alignment of con-ikot-ikots sequenced from the venom gland of *C. geographus* (Cikot-Cg1 - Cg7; GenBank accession number GAJN01000024-GAJN01000030). Multiple sequence alignment was performed using MAFFT auto alignment (version 7) (42). Cikot-Cs from *C. striatus* and Cikot-Cp from *C. purpurascens* (also known as p21a) are shown for comparison. Black arrow shows processed Cikot from *C. striatus*. Mature, active Cikot-Cs is a dimer of a covalent dimer (tetramer) (29). Mature Cikot-Cp likely forms a non-covalent dimer and was found with proline as well as hydroxyproline residues in two positions (white arrows) and an amidated His at the C terminus (30). Tryptic peptides identified by mass spectrometry for *C. geographus* Cikots are boxed. Number of cysteine residues (shown in bold) for sequences without signal peptides are provided next to sequences. Signal peptides are underlined. Amino acid conservations are denoted by an asterisk (*). Full stops (.) and colons (:) represent a low and high degree of similarity, respectively.

normalized gel spot volumes (normalized across the three fluorescent dyes and across the replicate gels) belonging to the same protein or protein isoform were combined and compared with reads per kilobase per million mapped reads (RPKM). Pooling of normalized gel volumes led to large standard errors, as spots with high interspecimen variation (*t* test $p > 0.05$, BVA analysis) were included in this analysis. However, despite these variations, gene and protein expression correlated well for most proteins of interest (Fig. 9). This is particularly the case for astacin-like 1 isoforms, conopressins,

conoporins, con-ikot-ikots and UP-Cg1 (Fig. 9). Correlation analysis was not performed for astacin-like 2 isoforms, as the encoding sequence could not be retrieved from the *C. geographus* venom gland transcriptome (tryptic peptides matched to a transcript from *C. bullatus*). Less correlation was apparent for conodipines, UP-Cg2 and UCRP-Cg1. It is now well established that gene expression is not always predictive of protein abundance with many other factors affecting protein abundances (34, 35). Several experimental limitations may also explain the lack of correlation observed; Mass spec-

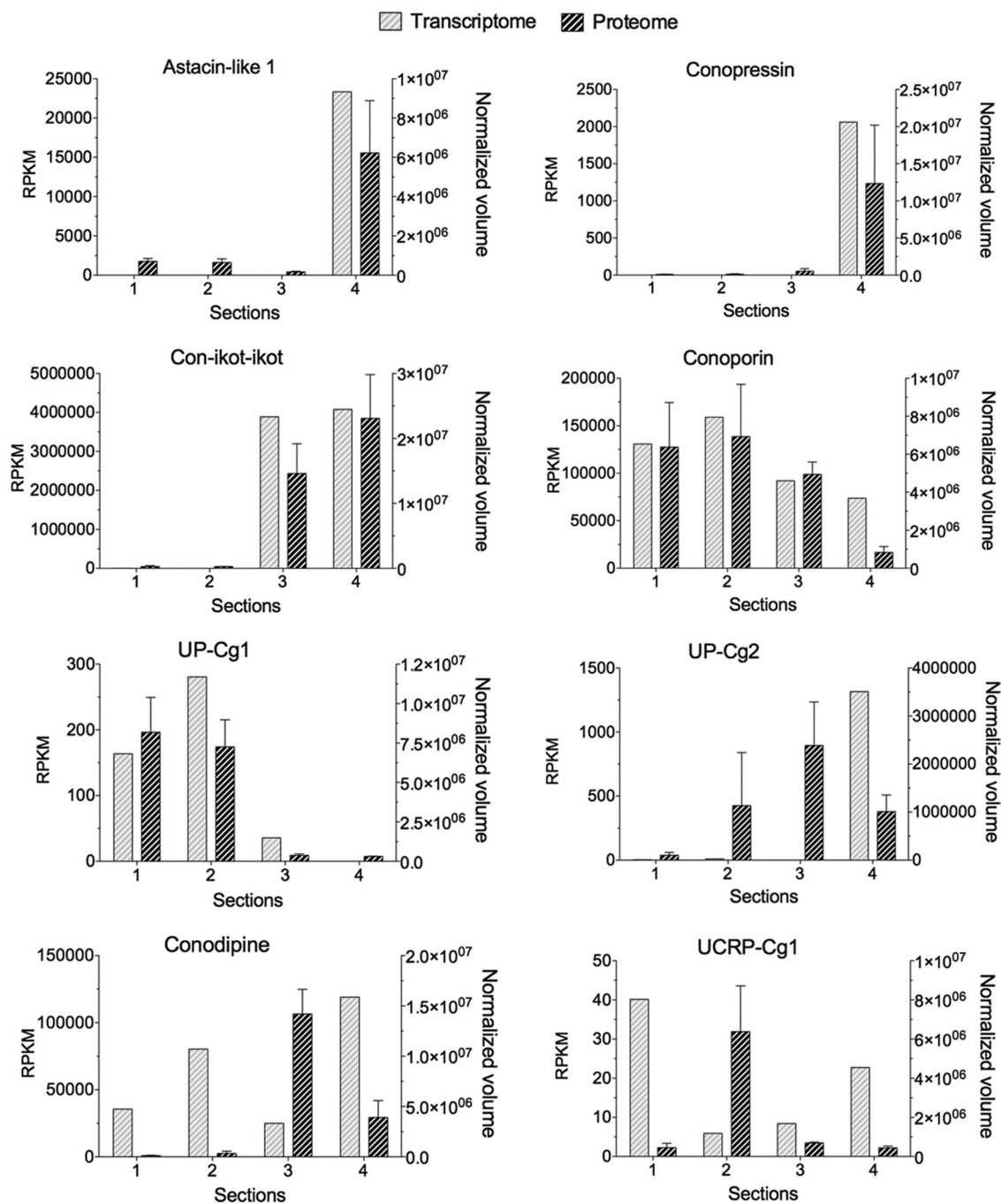


FIG. 9. Correlation of gene expression and protein abundance of selected proteins expressed in the four sections of the venom gland of *C. geographus*. Gene expression values are expressed as reads per kilobase per million mapped reads (RPKM) and shown on the left y axis. Protein abundance is represented by the normalized volume of the sum of all gel spots identified for a particular protein (right y axis). Normalized volumes are means (+ S.E.) from three specimens. Transcriptome analysis was performed on pooled samples from 4 specimens. All gene transcripts but UCRP-Cg1 showed significant differences in regional expression levels ($p < 0.01$; UCRP-Cg1: $p = 0.078$). Values are provided in supplemental File S2.

trometric analysis may have misidentified a protein spot because of co-migration of proteins on 2D gels. Additionally, transcriptome interrogation was performed on pooled RNA from four individuals, thus, not accounting for large interspecimen variations.

Notably, proteins important for conotoxin folding that appeared to be differentially expressed by DIGE but could not be statistically analyzed because of limitations with spot resolution exhibited significant regional changes in gene expression. These include different isoforms of PDI, PDIA6-an additional

TABLE II

Proteases identified in the venom gland of *Conus geographus* by transcriptome sequencing sorted after number of reads. Uniprot accession numbers of most homologous proteins are provided

Accession	Most homologous protein	Species	Reads	Domains
EKC20550	Tripeptidyl-peptidase 2	<i>Crassostrea gigas</i>	474	Peptidase S8
EKC26094	Astacin like-1 (Blastula protease)	<i>Crassostrea gigas</i>	145	M12A
Q9D958	Signal peptidase complex subunit 1	<i>Mus musculus</i>	68	SPC12
Q5BJI9	Signal peptidase complex subunit 2	<i>Danio rerio</i>	28	SPC25
Q3SZ71	Mitochondrial-processing peptidase	<i>Bos taurus</i>	24	M16
Q9VCA9	Signal peptidase complex subunit 3	<i>Drosophila melanogaster</i>	21	SPC22
P42893	Endothelin-converting enzyme 1	<i>Rattus norvegicus</i>	20	M13
P42674	Blastula protease 10	<i>Paracentrotus lividus</i>	16	M12A, MMP
P53634	Dipeptidyl-peptidase 1	<i>Homo sapiens</i>	13	Cathepsin C
P09648	Cathepsin L1	<i>Gallus gallus</i>	11	Peptidase C1A
P50579	Methionine aminopeptidase 2	<i>Homo sapiens</i>	10	APP-MetAP
EKC19359	Virulence metalloprotease	<i>Crassostrea gigas</i>	8	M4
P28840	Neuroendocrine convertase 1	<i>Rattus norvegicus</i>	7	S8, S53
Q32LQ0	Glutamyl aminopeptidase	<i>Bos taurus</i>	7	DUF
Q10714	Angiotensin-converting enzyme	<i>Drosophila melanogaster</i>	7	M2
P16519	Neuroendocrine convertase 2	<i>Homo sapiens</i>	5	Convertase P
P78562	Phosphate-regulating endopeptidase	<i>Homo sapiens</i>	4	M13
P13277	Digestive cysteine proteinase 1	<i>Homarus americanus</i>	4	Peptidase C1A
Q03168	Lysosomal aspartic protease	<i>Aedes aegypti</i>	4	Cathepsin D
Q4VC17	Metalloproteinase	<i>Mus musculus</i>	3	M12B
Q26563	Cathepsin C	<i>Schistosoma mansoni</i>	3	C1 Peptidase
P55110	Zinc metalloproteinase mspA	<i>Legionella longbeachae</i>	3	GluZincin
XP_002738143	Caseinolytic peptidase	<i>Saccoglossus kowalevskii</i>	3	ClpP
P37892	Carboxypeptidase E	<i>Lophius americanus</i>	3	M14
Q80W54	CAAX prenyl protease 1 homolog	<i>Mus musculus</i>	2	M48
Q8NDH3	Aminopeptidase NPEPL1	<i>Homo sapiens</i>	2	M17
A6H611	Mitochondrial intermediate peptidase	<i>Mus musculus</i>	2	M3
Q20176	Zinc metalloproteinase nas-39	<i>Caenorhabditis elegans</i>	2	ZnMc
A6QPT7	ER aminopeptidase 2	<i>Bos taurus</i>	2	M1
Q80Z60	Endothelin-converting enzyme 2	<i>Mus musculus</i>	2	M13
Q22523	Zinc metalloproteinase T16A9.4	<i>Caenorhabditis elegans</i>	2	M13
Q9R014	Cathepsin J	<i>Mus musculus</i>	2	C1 Peptidase
Q26534	Cathepsin L	<i>Schistosoma mansoni</i>	2	Peptidase C1A
P70669	Metalloendopeptidase homolog PEX	<i>Mus musculus</i>	2	M13
P79171	Aminopeptidase N	<i>Felis catus</i>	1	DUF
Q0VGW4	ER metallopeptidase 1	<i>Xenopus laevis</i>	1	M28
Q5G269	Neurotrypsin	<i>Pongo pygmaeus</i>	1	Tryp-SPc
P12955	Xaa-Pro dipeptidase	<i>Homo sapiens</i>	1	APP-MetAP
O75976	Carboxypeptidase D	<i>Homo sapiens</i>	1	M14 N/E
P05689	Cathepsin Z	<i>Bos taurus</i>	1	C1 Peptidase
Q5R432	Cytosolic non-specific dipeptidase	<i>Pongo abelii</i>	1	M20
P42658	Dipeptidyl aminopeptidase-like 6	<i>Homo sapiens</i>	1	N/A
P0C1T0	Membrane metallo-endopeptidase-like 1	<i>Rattus norvegicus</i>	1	M13
Q9H3G5	Serine carboxypeptidase	<i>Homo sapiens</i>	1	S10
Q9N2V2	Zinc metalloproteinase nas-30	<i>Caenorhabditis elegans</i>	1	M12A
Q4LAL9	Cathepsin D	<i>Canis familiaris</i>	1	Cathepsin D
EKC42374	Aminopeptidase N	<i>Crassostrea gigas</i>	1	M1
Q5XIB4	Ufm1-specific protease 2	<i>Rattus norvegicus</i>	1	N/A
Q2HJH1	Aspartyl aminopeptidase	<i>Bos taurus</i>	1	M18, M20, M28, M42
Q2KJ83	Carboxypeptidase N	<i>Bos taurus</i>	1	M14
Q35186	Cathepsin K	<i>Rattus norvegicus</i>	1	C1 Peptidase
P00747	Plasminogen	<i>Homo sapiens</i>	1	KR
A6NEC2	Puromycin-sensitive aminopeptidase	<i>Homo sapiens</i>	1	M1
P59110	Sentrin-specific protease 1	<i>Mus musculus</i>	1	N/A

member of the PDI family- and P4H (supplemental File S2). Similar to observations made by DIGE (see group 3, Fig. 1) most of these transcripts have highest expression levels in section 3 and 4.

Transcriptome analysis further revealed a yet unknown genetic diversity of venom gland proteases (Table II). Because of a low number of reads only 2 of the 54 proteases identified exhibited significant regional expression patterns across the

gland. Tripeptidyl-peptidase 2 had the highest number of reads in section 2 (419,240 out of 704,845 RPKM). Correlation with proteome data was not performed as mass spectrometric analysis did not identify this protein in the venom gland. The second protease with significant changes in expression was astacin-like 1, which was shown to be differentially expressed by DIGE analysis and correlate well with gene expression (Fig. 9).

DISCUSSION

The present study utilized a combined transcriptomic and proteomic approach to investigate regional protein abundance and gene expression profiles along the heterogeneous venom gland of *C. geographus*. Transcriptome sequencing provided a database for mass spectrometric protein identifications at the same time informing on regional differences in gene expression. Combined with the resolution power of 2D-DIGE this methodology revealed a yet unknown diversity of proteins and gene transcripts and distinct expression patterns of several highly abundant proteins and their isoforms. These findings clearly demonstrate that the venom gland of cone snails is compartmentalized into functionally distinct entities.

Most prominent changes in protein abundance were observed between most distant regions with comparatively few changes in proximal segments close to the bulb. This is consistent with overall gene expression profiles demonstrating greatest correlation between the proximal sections 1 and 2 with only little correlation between distant regions (7).

Our findings are in agreement with and complement previous studies on regional changes in conotoxin gene expression and abundance along the venom glands of *C. geographus* (7) and *C. textile* (8, 36). For example, in *C. textile*, a mollusk-hunting species with well characterized venom, conotoxins of the M-superfamily are more abundant in the proximal region of the gland whereas O-superfamily peptides are predominantly found in the central-distal part (6). In *C. geographus* O-superfamily toxins are expressed in all four sections whereas T-superfamily peptides are mostly found toward the pharynx (7).

Although more studies are needed to elucidate the specialized regional adaptations of glandular epithelial cells for the biosynthesis, modification, storage and secretion of a particular venom component, it is likely that these cells harbor different sets of “helper” proteins that are important for the proper assembly and transport of venom peptides/proteins. These include molecular chaperones of the heat shock protein family (17), different isoforms of PDI (17, 18) and modifying enzymes such as P4H. Transcriptome analysis showed differential expression of these helper proteins, however, because of technical limitations these findings could only be partially confirmed by DIGE. Future studies utilizing larger 2D gels for better protein spot resolution and additional full-length sequences of protein isoforms are likely to verify these initial observations.

Cone snail venom is among the most diverse venoms found in the animal kingdom. Recent studies have suggested that, at any given time, thousands of different toxin peptides are biosynthesized, modified and secreted by the epithelial cells of the venom gland (2). *C. geographus* transcriptome sequencing revealed that conotoxin transcripts account for 88% of all aligned reads within annotated transcripts with more species-specific gene products than described for any other tissue known to date (7). We propose that the compartmentalization of the *Conus* venom gland described here is a key evolutionary innovation for the high-throughput production of such a high density and diversity of venom compounds.

The evolutionary origin of the venom gland has long been subject to speculation. However, recent morphological evidence obtained on different life stages of *Conus lividus* strongly suggests that the venom gland evolved by rapid pinching-off from the epithelium of the mid-esophageal wall (37). The epithelial tubes and vesicles of the venom gland thus originated from a pre-existing epithelial sheet of the mid-esophagus (37). This epithelial remodeling process gave rise to gland elongation, epithelial cell specialization and the formation of the muscular venom bulb. This evolutionary relationship between the venom gland and the esophagus is likely to be reflected in an overlap in the transcriptome and proteome between these two tissue types. Although this has not been addressed yet, several studies have reported similarities between the venom gland and the salivary gland (27, 38).

Interestingly, among the transcripts that were more abundant in the salivary gland of *C. consors* when compared with the venom gland were conoporins (27), highly expressed proteins in the venom gland of *C. geographus* identified herein. Conoporins share homology with actinoporins and echotoxins, potent cytolytic and hemolytic proteins that exert toxicity through forming oligomeric cation-selective pores in membranes leading to osmotic shock and cell death (19). The presence of a homologous protein in the dissected and injected venom of *C. consors* is a recent observation that was facilitated by the availability of next generation sequencing data (27, 39). Here, nine conoporin transcripts were identified uncovering a yet unknown genetic diversity of this venom component. Remarkably, the same applies to other putative venom components of high abundance and distinct regional expression patterns, including conodipines, conopressins, and con-ikot-ikots. What appears to be functional redundancy of venom components is often subsequently identified as generation of compounds with subtype specificities for their target receptors and ion channel. Whether the various protein isoforms identified here evolved to distinguish between different target subtypes warrants further investigation.

In addition to known venom components, this study identified a number of proteins of yet unknown function. Notably UP-Cg1 and its isoforms were among the most highly abundant proteins in the proximal part of the gland and likely serve an important role in venom biosynthesis or envenomation. A

recent survey of the venom gland of *C. consors* also described several unknown proteins (27) with no apparent overlap between proteins identified here. This finding strongly points toward the evolution of species-specific compounds that reflect the ecological niche of a particular species. Future recombinant expression and characterization of these unknown compounds is likely to identify their functional role in the *Conus* venom gland and may provide novel reagents for pharmacotherapeutic studies.

Transcriptome sequencing further revealed the presence of a great diversity of proteases with various functional domains including zinc-binding, cathepsin-like, serine protease and peptidase domains. The functional role of proteases in the venom gland is not easily determined. Most conotoxins are translated as precursor proteins with an N-terminal signal sequence followed by a propeptide that is proteolytically cleaved during toxin maturation. Cleavage occurs at a pair of basic residues, however many conotoxins contain variations of this site (e.g. ER, TK, TR) suggesting that conotoxin cleavage requires a diverse set of proteases with different substrate specificities. The only protease proposed to play a role in propeptide cleavage is Tex31, a member of the cysteine-rich secretory protein (CRISP) family originally isolated from the venom of *C. textile* (40). Although recombinant Tex31 proteolytically processed a number of conotoxin-like peptides *in vitro* controversy exists on the real role of this protein in the venom. Given its homology to proteases utilized by other venomous animals to cause tissue disruption, Tex31 was suggested to function in envenomation rather than conotoxin processing (41). Transcriptomic and proteomic analysis did not identify Tex31 in the venom gland of *C. geographus* but revealed high abundances of 2 zinc metalloproteases, astacin-like 1 and 2. The role of these proteases remains to be functionally assessed, however, recent evidence on the direct interaction between *Conus* astacin-like 1 with a conotoxin propeptide indicates a role of this protease in the proteolytic processing of conotoxin precursors (Safavi-Hemami, unpublished data). Highest gene expression and protein abundance of this protease was observed in section 4 where activation of conotoxins may be a crucial step preceding envenomation.

By combining transcriptomic and proteomic methodologies this study shed new light on the genetic and proteomic diversity of venom gland components and provided molecular and biochemical evidence for the compartmentalization of this tissue into functional entities. We propose that functionalization of the *Conus* venom gland is a key evolutionary innovation for the high-throughput production of venom polypeptides.

Acknowledgments—We thank Dr. David Perkins and Paul O'Donnell for assistance with database submissions.

*This work was partially supported by a Discovery Grant (DP110101331) from the Australian Research Council (BMO, AWP, and a program project grant (GM48677) from the National Institute of

General Medical Sciences (PB, BMO) and R01GM099939 (PB, MY). AWP acknowledges fellowship support from the Australian National Health and Medical Research Council. HSH is supported by a Marie Curie Fellowship from the European Union (CONBIOS 330486).

§ This article contains supplemental Files S1 and S2, Figs. S1 and S2, and Table S1.

‡‡ To whom correspondence should be addressed: Department of Biology, 257 South 1400 East, University of Utah, Salt Lake City, UT 84112, USA. Tel.: +1 801 581 8370; Fax: +1 801 585 5010; E-mail: safavihelena@gmail.com.

REFERENCES

1. Livett, B. G., Sandall, D., Keays, D., Down, J. G., Gayler, K., Satkunanathan, N., and Khalil, Z. (2006) Therapeutic applications of conotoxins that target the neuronal nicotinic acetylcholine receptor. *Toxicon* **48**, 810–829
2. Davis, J., Jones, A., and Lewis, R. J. (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* **30**, 1222–1227
3. Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E. E., Woodward, S. R., Hillyard, D. R., and Cruz, L. J. (1990) Diversity of *Conus* neuropeptides. *Science* **249**, 257–263
4. Marshall, J., Kelley, W. P., Rubakhin, S. S., Bingham, J.-P., Sweedler, J. V., and Gilly, W. F. (2002) Anatomical Correlates of Venom Production in *Conus californicus*. *Biol. Bull.* **203**, 27–41
5. Safavi-Hemami, H., Young, N. D., Williamson, N. A., and Purcell, A. W. (2010) Proteomic interrogation of venom delivery in marine cone snails - Novel insights into the role of the venom bulb. *J. Proteome Res.* **9**, 5610–5619
6. Dobson, R., Collodoro, M., Gilles, N., Turtoi, A., De Pauw, E., and Quinton, L. (2012) Secretion and maturation of conotoxins in the venom ducts of *Conus textile*. *Toxicon* **60**, 1370–1379
7. Hu, H., Bandyopadhyay, P. K., Olivera, B. M., and Yandell, M. (2012) Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics* **13**, 1471-2164-13-284
8. Tayo, L. L., Lu, B. W., Cruz, L. J., and Yates, J. R. (2010) Proteomic Analysis Provides Insights on Venom Processing in *Conus textile*. *J. Proteome Res.* **9**, 2292–2301
9. Veith, P. D., O'Brien-Simpson, N. M., Tan, Y., Djatmiko, D. C., Dashper, S. G., and Reynolds, E. C. (2009) Outer membrane proteome and antigens of *Tannerella forsythia*. *J. Proteome Res.* **8**, 4279–4292
10. Hu, H., Bandyopadhyay, P. K., Olivera, B. M., and Yandell, M. (2011) Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics* **12**, 60
11. Vizcaino, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069
12. Chevreaux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E., Wetter, T., and Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–1159
13. Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760
14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
15. Zdobnov, E. M., and Apweiler, R. (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848
16. Robinson, M. D., and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25
17. Safavi-Hemami, H., Gorasia, D. G., Steiner, A. M., Williamson, N. A., Karas, J. A., Gajewiak, J., Olivera, B. M., Bulaj, G., and Purcell, A. W. (2012) Modulation of conotoxin structure and function is achieved through a multienzyme complex in the venom glands of cone snails. *J. Biol. Chem.* **287**, 34288–34303
18. Wang, Z. Q., Han, Y. H., Shao, X. X., Chi, C. W., and Guo, Z. Y. (2007) Molecular cloning, expression and characterization of protein disulfide

- isomerase from *Conus marmoreus*. *FEBS J.* **274**, 4778–4787
19. Alegre-Cebollada, J., Oñaderra, M., Gavilanes, J. G., and del Pozo, A. M. (2007) Sea anemone actinoporins: the transition from a folded soluble state to a functionally active membrane-bound oligomeric pore. *Current Protein Peptide Sci.* **8**, 558–572
 20. Castañeda, O., Sotolongo, V., Amor, A. M., Stöcklin, R., Anderson, A. J., Harvey, A. L., Engström, A., Wernstedt, C., and Karlsson, E. (1995) Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*. *Toxicon* **33**, 603–613
 21. Rangaraju, S., Khoo, K. K., Feng, Z. P., Crossley, G., Nugent, D., Khaytin, I., Chi, V., Pham, C., Calabresi, P., Pennington, M. W., Norton, R. S., and Chandy, K. G. (2010) Potassium channel modulation by a toxin domain in matrix metalloprotease 23. *J. Biol. Chem.* **285**, 9124–9136
 22. Sachs, H., and Takabatake, Y. (1964) Evidence for a precursor in vasopressin biosynthesis. *Endocrinology* **75**, 943–948
 23. Cruz, L. J., de Santos, V., Zafaralla, G. C., Ramilo, C. A., Zeikus, R., Gray, W. R., and Olivera, B. M. (1987) Invertebrate vasopressin/oxytocin homologs. Characterization of peptides from *Conus geographus* and *Conus striatus* venoms. *J. Biol. Chem.* **262**, 15821–15824
 24. Nielsen, D. B., Dykert, J., Rivier, J. E., and McIntosh, J. M. (1994) Isolation of Lys-conopressin-G from the venom of the worm-hunting snail, *Conus imperialis*. *Toxicon* **32**, 845–848
 25. Lirazan, M., Jimenez, E. C., Craig, A. G., Olivera, B. M., and Cruz, L. J. (2002) Conophysin-R, a *Conus radiatus* venom peptide belonging to the neurophysin family. *Toxicon* **40**, 901–908
 26. McIntosh, J. M., Ghomashchi, F., Gelb, M. H., Dooley, D. J., Stoehr, S. J., Giordani, A. B., Naisbitt, S. R., and Olivera, B. M. (1995) Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*. *J. Biol. Chem.* **270**, 3518–3526
 27. Leonardi, A., Biass, D., Kordiš, D., Stöcklin, R., Favreau, P., and Križaj, I. (2012) *Conus* consors snail venom proteomics proposes functions, pathways, and novel families involved in its venom system. *J. Proteome Res.* **11**, 5046–5058
 28. Tsai, I. H., Lu, P. J., Wang, Y. M., Ho, C. L., and Liaw, L. L. (1995) Molecular cloning and characterization of a neurotoxic phospholipase A2 from the venom of Taiwan habu (*Trimeresurus mucrosquamatus*). *Biochem. J.* **311**, 895–900
 29. Walker, C. S., Jensen, S., Ellison, M., Matta, J. A., Lee, W. Y., Imperial, J. S., Duclos, N., Brockie, P. J., Madsen, D. M., Isaac, J. T. R., Olivera, B., and Maricq, A. V. (2009) A novel *Conus* snail polypeptide causes excitotoxicity by blocking desensitization of AMPA receptors. *Curr. Biol.* **19**, 900–908
 30. Möller, C., and Marí, F. (2011) 9.3 kDa components of the injected venom of *Conus purpurascens* define a new five-disulfide conotoxin framework. *Biopolymers* **96**, 158–165
 31. Violette, A., Leonardi, A., Piquemal, D., Terrat, Y., Biass, D., Dutertre, S., Noguier, F., Ducancel, F., Stöcklin, R., Križaj, I., and Favreau, P. (2012) Recruitment of glycosyl hydrolase proteins in a cone snail venomous arsenal: further insights into biomolecular features of *Conus* venoms. *Marine Drugs* **10**, 258–280
 32. Ritonja, A., Meloun, B., and Gubensek, F. (1983) The primary structure of *Vipera ammodytes* venom trypsin inhibitor I. *Biochim. Biophys. Acta* **748**, 429–435
 33. Francischetti, I. M., My-Pham, V., Harrison, J., Garfield, M. K., and Ribeiro, J. M. (2004) *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* **337**, 55–69
 34. Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M., and Nilsson, P. (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* **10**, 1471–2164–10–365
 35. Vogel, C., and Marcotte, E. M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**, 227–232
 36. de Plater, G. M., Martin, R. L., and Milburn, P. J. (1998) A C-type natriuretic peptide from the venom of the platypus (*Ornithorhynchus anatinus*): structure and pharmacology. *Comp. Biochem. Physiol.* **120**, 99–110
 37. Page, L. R. (2012) Developmental modularity and phenotypic novelty within a biphasic life cycle: morphogenesis of a cone snail venom gland. *Proc. Biol. Sci.* **270**, 77–83
 38. Biggs, J. S., Olivera, B. M., and Kantor, Y. I. (2008) Alpha-conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon* **52**, 101–105
 39. Violette, A., Biass, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F., Stöcklin, R., and Favreau, P. (2012) Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J. Proteomics* **75**, 5215–5225
 40. Milne, T. J., Abbenante, G., Tyndall, J. D. A., Halliday, J., and Lewis, R. J. (2003) Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J. Biol. Chem.* **278**, 31105–31110
 41. Qian, J., Guo, Z. Y., and Chi, C. W. (2008) Cloning and isolation of a *Conus* cysteine-rich protein homologous to Tex31 but without proteolytic activity. *Acta Biochim. Biophys. Sinica* **40**, 174–181
 42. Katoh, K., and Standley, D. M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evolution* **30**, 772–780
 43. Zdobnov, E. M., and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848