JOURNAL OF PROTEOMICS 102 (2014) 1–10

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Multifunctional warheads: Diversification of the toxin arsenal of centipedes via novel multidomain transcripts

Eivind A.B. Undheim[a,b], *Kartik Sunagar*[c,d], *Brett R. Hamilton*[e], *Alun Jones*[a], *Deon J. Venter*[e,f], *Bryan G. Fry*[a,b,*], *Glenn F. King*[a,**]

[a]*Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia*
[b]*School of Biological Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia*
[c]*Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal*
[d]*CIIMAR/CIMAR–Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas 289, P 4050-123 Porto, Portugal*
[e]*OMICS, Pathology Department, Mater Health Services, South Brisbane, QLD 4101, Australia*
[f]*School of Medicine, The University of Queensland, St. Lucia, QLD 4072, Australia*

ARTICLE INFO

Article history:
Received 5 January 2014
Accepted 21 February 2014
Available online 3 March 2014

Keywords:
Centipede venom
Multidomain transcript
Evolution
Posttranslational modification
MALDI imaging

ABSTRACT

Arthropod toxins are almost invariably encoded by transcripts encoding prepropeptides that are posttranslationally processed to yield a single mature toxin. In striking contrast to this paradigm, we used a complementary transcriptomic, proteomic and MALDI-imaging approach to identify four classes of multidomain centipede-toxin transcripts that each encodes multiple mature toxins. These multifunctional warheads comprise either: (1) repeats of linear peptides; (2) linear peptides preceding cysteine-rich peptides; (3) cysteine-rich peptides preceding linear peptides; or (4) repeats of linear peptides preceding cysteine-rich peptides. MALDI imaging of centipede venom glands revealed that these peptides are posttranslationally liberated from the original gene product in the venom gland and not by proteases following venom secretion. These multidomain transcripts exhibit a remarkable conservation of coding sequences, in striking contrast to monodomain toxin transcripts from related centipede species, and we demonstrate that they represent a rare class of predatory toxins that have evolved under strong negative selection. We hypothesize that the peptide toxins liberated from multidomain precursors might have synergistic modes of action, thereby allowing negative selection to dominate as the toxins encoded by the same transcript become increasingly interdependent.

Biological significance
These results have direct implications for understanding the evolution of centipede venoms, and highlight the importance of taking a multidisciplinary approach for the investigation of novel venoms. The potential synergistic actions of the mature peptides are also of relevance to the growing biodiscovery efforts aimed at centipede venom. We also demonstrate the application of

* *Correspondence to:* B.G. Fry, School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia. Tel.: +61 400 1931 832.
** *Correspondence to:* G.F. King, Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia, QLD 4072, Australia. Tel.: +61 7 3346 2025; fax: +61 7 3346 2021.
E-mail addresses: bgfry@uq.edu.au (B.G. Fry), glenn.king@imb.uq.edu.au (G.F. King).

http://dx.doi.org/10.1016/j.jprot.2014.02.024

MALDI imaging in providing a greater understanding of toxin production in venom glands. This is the first MALDI imaging data of any venom gland.

1. Introduction

Proteins earmarked for secretion are typically produced as prepropeptides comprised of a signal peptide and one or two propeptide regions that are posttranslationally excised by endoproteases to yield a single mature protein product [1–3]. There are, however, a number of deviations from this scheme, including transcripts that lack propeptide-encoding regions and transcripts that encode multiple mature proteins [3]. For example, in both vertebrates and invertebrates, neuropeptides and hormones are commonly produced as multifunctional precursors containing a signal peptide and multiple copies of the neuropeptide or hormone separated by propeptide regions [4].

Multifunctional toxin transcripts, however, are exceedingly rare in most venomous taxa, with the reptilian clade Toxicofera being the notable exception. Various strategies leading to multifunctional toxin transcripts have evolved both convergently and divergently on several occasions within Toxicofera. These include the duplication events leading to precursors encoding tandem stretches of sarafotoxins in *Atractaspis* snakes [5], seven newly evolved bradykinin potentiating peptides in the propeptide region of the precursor encoding a C-type natriuretic peptide in the pit viper *Bothrops jararaca* [6], and multiple helokinestatin peptides in the propeptide region of the precursor encoding a B-type natriuretic peptide in Helodermatidae and Anguiidae lizard venoms [7–9]. The venom glands of coleoid cephalopods (cuttlefish, octopus and squid) also produce multifunctional transcripts encoding 3–4 pacifastin peptides that are posttranslationally liberated [10].

In striking contrast to coleoids and toxicoferans, invertebrate venomous animals such as marine cone snails, hymenopterans, sea anemones, scorpions, and spiders strictly adhere to the canonical one gene–one toxin strategy [11,12]. The most widely studied venomous arthropods, namely spiders and scorpions, generate venom diversity via expression of numerous isoforms of each toxin type rather than via multiple posttranslational modifications of a single translated product [12,13]. Their impressive toxin arsenal [14] appears to have evolved through classical gene duplication events followed by explosive diversification driven by positive selection [12,15]. In contrast, transcripts encoding multiple mature toxins are extremely rare in arthropods, and have only been noted for latarcins, linear antimicrobial peptides found in the venom of the spider *Lachesana tarabaevi* [16].

Centipedes may be the oldest extant terrestrial venomous lineage, having arisen more than 400 million years ago (Mya) [17]. Reflecting this ancient divergence, the centipede venom apparatus as well as most centipede toxins described to date bear little resemblance to those of other arthropods [18–20]. However, the centipede toxin transcripts described to date conform to the arthropod paradigm of encoding a prepropeptide containing a single mature toxin domain. In striking contrast, we describe here four different types of multidomain transcripts from the venom gland of four species of scolopendrid centipede and use MALDI imaging to show that these multifunctional "warheads" are activated in the venom gland prior to venom expulsion.

2. Material and methods

2.1. Specimen and venom collection

Ethmostigmus rubripes was purchased from Mini Beast Wildlife (www.minibeastwildlife.com.au), *Scolopendra morsitans* was collected from the Darling Downs region, Queensland, Australia, and *Cormocephalus westwoodi* was collected from the Launceston region, Tasmania, Australia; all were identified according to Koch [21–23]. *Scolopendra alternans* (Haiti) were purchased from La Ferme Tropicale (www.lafermetropicale.com). For venom collection, centipedes were starved for 3 weeks, then anesthetized with CO_2 and venom extracted by electrostimulation (12 V, 1 mA). All species were milked except *S. alternans*. Venom was immediately lyophilised and stored until further use at −80 °C.

2.2. cDNA library construction

Four days after venom depletion by electrostimulation, the venom glands were removed from five anesthetized specimens, flash frozen, and pooled. Total RNA was extracted by using TRIzol (Life Technologies) and enriched for mRNA using a DynaBeads Direct mRNA kit (Life Technologies). mRNA was reverse transcribed, fragmented and ligated to a unique 10-base multiplex identifier (MID) tag and applied to a PicoTitrePlate for simultaneous amplification and sequencing on a Roche 454 GS FLX+ Titanium platform (Australian Genome Research Facility). Automated grouping and analysis of sample-specific MID reads enabled informatic separation of sequences from other transcriptomes on the plates. Low-quality sequences were removed prior to de novo contig assembly using MIRA (version 3.4.0.1). Assembly details (number of reads, average read length, number of contigs and average assembled bases per contig) were: *E.rubripes* 72740, 375, 6980, 1035; *C. westwoodi* 48041, 376, 1706, 544; *S. alternans* 57175, 355, 5044, 612; *S. morsitans* 93436, 356, 6029, 621. Contigs were processed and analyzed using CLC Main Work Bench (ver. 6.2; CLC bio) and the Blast2GO bioinformatic suite [24,25]. To identify putative toxin transcripts, each transcriptome was searched against the Tox-Prot database (http://www.uniprot.org/program/Toxins) to which additional functionally annotated centipede toxin sequences were added [18,20]; the results are shown in Supplementary Fig. 1. Data can be accessed at the National Center for Biotechnology Information under bioprojects PRJNA200639 (*E. rubripes*), PRJNA200641 (*C. westwoodi*), PRJNA200753 (*S. alternans*), and PRJNA200640 (*S. morsitans*).