Mitchell Michela (Orcid ID: 0000-0001-6331-534X)

Title: Evolution of cnidarian *trans*-defensins: sequence, structure and exploration of chemical space

Short Title: Evolution of cnidarian *trans*-defensins

Michela L. Mitchell[1,2,3,4#], Thomas Shafee[5,6#]*, Anthony T. Papenfuss[2,7,8,9,10], Raymond S. Norton[1]

[1] *Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Melbourne, Victoria, 3052, Australia*
[2] *Bioinformatics Division, Walter & Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia*
[3] *Museum Victoria, GPO Box 666, Melbourne, Vic, 3001, Australia*
[4] *Queensland Museum, PO Box 3300, South Brisbane, Queensland, 4101, Australia*
[5] *Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria, 3086, Australia*
[6] *Department of Animal, Plant, and Soil Sciences, AgriBio, La Trobe University, Melbourne, Victoria, 3086, Australia*
[7] *Peter MacCallum Cancer Centre, Melbourne, Victoria, 3000, Australia.*
[8] *Department of Medical Biology, University of Melbourne, Melbourne, Victoria 3010, Australia.*
[9] *Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria 3010, Australia.*
[10] *Department of Mathematics and Statistics, University of Melbourne, Melbourne, Victoria 3010, Australia.*

[#] *Co-first author*
* *Correspondence to: T.Shafee@LaTrobe.edu.au*

Keywords: trans-*defensins, Cnidaria, sea anemone, Corallimorpharia, evolution, cysteine-rich peptides*

Abstract

Many of the small, cysteine-rich ion-channel modulatory peptides found in Cnidaria are distantly related to vertebrate defensins (of the *trans*-defensin superfamily). Transcriptomic and proteomic studies of the endemic Australian speckled sea anemone (*Oulactis* sp.) yielded homologous peptides to known defensin sequences. We extended these data using existing and custom-built hidden Markov models to extract defensin-like families from the transcriptomes of seven endemic Australian cnidarian species. Newly sequenced transcriptomes include three species of Actiniaria (true sea anemones); the speckled anemone (*Oulactis* sp.), *Oulactis muscosa*, *Dofleinia* cf. *armata* and a species of Corallimorpharia, *Rhodactis* sp. We analysed these novel defensin-like sequences along with published homologues to study the evolution of their physico-chemical properties in vertebrate and invertebrate fauna. The cnidarian *trans*-defensins form a distinct cluster within the chemical space of the superfamily, with a unique set of motifs and biophysical properties. This cluster contains identifiable subgroups, whose distribution in the space also correlates with the divergent evolution of their structures. These sequences, currently restricted to cnidarians, form an evolutionarily distinct clade within the *trans*-defensin superfamily.

1. Introduction

Defensins are short proteins that can be classified into two superfamilies (*cis*- and *trans*-) with independent evolutionary origins. Together, these superfamilies encompass members with a range of biological activities including antimicrobial peptides (AMPs), neurotoxins, and enzyme inhibitors (Shafee, Lay, Phan, Anderson, & Hulett, 2017). *Trans*-defensins are found in vertebrate animals (including α-, β-, and θ-defensins) and some invertebrate phyla (Zhu & Gao, 2013), whereas *cis*-defensins are found in plants, fungi and several other invertebrate phyla (Shafee et al., 2017). Cnidaria are unusual in that they contain sequences from both superfamilies (Shafee, Lay, Hulett, & Anderson, 2016; Shafee et al., 2017):, for example antimicrobial hydramacins from the *cis*-defensins, and enzyme inhibitors and toxins from the *trans*-defensins. We therefore focus our search for additional ion-channel modulators on the *trans*-defensins.

Although most members of the *trans*-defensin superfamily that have been functionally characterised are antimicrobial, a number of defensin-like proteins (DLPs) have been recruited to alternative biological functions. Toxic peptides based on this fold are found within the venom of platypus (ovDLP), snakes (e.g. crotamine), bearded lizards (e.g. helofensin), and sea anemones (e.g. APETx2) (Chagot, Diochot, Pimentel, Lazdunski, & Darbon, 2005; Coronado et al., 2013; Fry et al., 2010; Roelants et al., 2010; Torres et al., 1999). The fold has similarly been recruited as enzyme inhibitors in ticks (e.g. TCI) and a sea anemone (helianthamide) (Arolas et al., 2005; Tysoe et al., 2016).

AMPs and defensin sequences have been isolated from the venoms of coral (*Pocillopora damicornis* – damicornin), hydroids (*Hydra vulgaris* – Hydramacin-1) and sea anemones (*Anemonia sulcata* (syn. *A. viridis*) – BDS-I) (Driscoll, Gronenborn, Béress, & Clore, 1989; Jung et al., 2009; Vidal-Dupiol et al., 2011). Recently, a defensin was isolated from the venom of *Urticina eques* (τ-AnmTx Ueq 12-1) (Logashina et al., 2017) that has a dual functionality, displaying both antibacterial activity and potentiating activity against the transient receptor potential ankyrin 1 (TRPA1) (Logashina et al., 2017). Defensins do not reside only in venom; a novel β-defensin α-amylase inhibitor has been isolated from the mucus of the sea anemone *Heteractis magnifica* (Sintsova et al., 2018).

Defensin-like peptides from cnidarians thus represent a potentially valuable class of bioactive peptides. Nonetheless, there is a paucity of sequences available in public databases. Since cnidarians form one of the oldest clades of venomous animals (~700 *mya*) and are prevalent throughout the world's oceans, this paucity of cnidarian data means that evolutionary studies commonly lack this most basal and widely ranging group of venomous animals.

In this study we examine cnidarian defensin sequences, and their chemical diversity in relation to the other invertebrate and vertebrate members of the *trans*-defensin superfamily. We extend a previous dataset of these sequences (Shafee, Lay, et al., 2016) by using hidden Markov models (HMMs) to search newly-sequenced cnidarian transcriptomes. Novel transcriptomes include three Australian sea anemones (*Oulactis* sp., *O. muscosa*, *Dofleinia* cf. *armata*) and a corallimorpharian (*Rhodactis* sp.). Additional published transcriptomes include *Nematostella vectensis*, *Edwardsiella lineata* and two clades of cnidarian symbiont algae *Symbiodinium* (Bayer et al., 2012; Lubinski et al., 2014; Stefanik et al., 2014).

The short length and low sequence identity of defensin sequences can make them challenging to study using traditional methods such as phylogenetics (Zielezinski, Vinga, Almeida, & Karlowski, 2017). Repeated mutation at the same location can obscure phylogenetic relationships by saturation effects and long branch attraction (Philippe et al., 2011). It has not been possible to resolve the phylogeny of the whole superfamily, although smaller phylogenies of the α-, β- and big-defensin families have been published previously (Zhu & Gao, 2013). Instead, we build a model of the superfamily's local region of sequence space, placing peptides into this space based on their biophysical properties (Shafee & Anderson, 2018). Since the 20 standard amino acids are each chemically distinct, the sequence space and chemical space are equivalent (map 1-to-1). We found that cnidarian sequences (with the exception of hydramacin) form a distinct cluster that lies between previously characterised vertebrate and invertebrate defensin groups, and appear to be an evolutionary distinct subclade. In addition, we observed a conserved set of residues and properties within the defensin scaffold that is restricted to cnidarians.

2. Methods

2.1 Cnidaria collection and aquaria

Specimens of *Oulactis* sp. were collected by hand from the intertidal zone, Brighton, Victoria, Australia, under Museum Victoria Permit number RP699. Specimens of *Rhodactis* sp. were collected by SCUBA, Stradbroke Island, Queensland. *Dofleinia* cf. *armata* was collected by SCUBA at the Gold coast seaway, Southport, Queensland. *Oulactis muscosa* was collected by hand from the intertidal zone, Elephant rock, Currumbin, Queensland. Queensland collecting was conducted under Queensland Museum permit 160782.

Animals were kept in aquaria at room temperature and fed prawn meat every three days. After tissue collection, specimens were preserved and deposited in the Museum Victoria (NMV) and Queensland museum (QM) invertebrate collections. *Oulactis* sp. NMV F223094, NMV F248416, NMV F248417, NMV F248418; *Rhodactis* sp. QM G335701, QM G335702;

Dofleinia cf. *armata* QM G335858, QM G335860; *Oulactis muscosa* QM G335853. All animals were identified by MLM.

2.2 Tissue collection & total RNA extraction

Animals were kept in aquaria for a minimum of one week prior to sampling and not fed for three days prior to tissue collection. Between 20 and 40 mg of various morphological tissue types was removed from each individual biological replicate and sent for sequencing as follows (number of replicates indicated in brackets): *Rhodactis* sp. – discal tentacle (1), marginal tentacle (1) and mesenterial filament (1); *Oulactis* sp. tentacles (2), acrorhagi/frill (1), mesenterial filament/gametes (1); *Oulactis muscosa* – tentacles (2), acrorhagi/frill (1); *Dofleinia* cf. *armata* – inner tentacles (2), outer tentacles (2), mesenterial filament/gametes (2). Tissue was immediately transferred to sterile tubes containing 600 µL of RNA*later*®. Samples were stored at 4 °C for two days prior to performing Total RNA extractions, except for *Rhodactis* sp. which was stored at -20 °C prior to tissue extraction after storage at 4 °C.

Tissue samples were lysed using a TissueLyser II (3 mm tungsten carbide beads). Total RNA extractions were performed using QIAGEN RNeasy Mini Kit standard and protocol, incorporating the optional DNA clean-up (QIAGEN RNase-Free DNase). Extractions were stored at -20 °C prior to being sent for quality testing (Agilent 2100 Bioanalyzer). *Oulactis* spp. and *Rhodactis* sp. samples selected for sequencing were transferred to RNAstable® tubes and dried overnight. *Dofleinia* cf. *armata* total RNA extractions were stored at -20 °C prior to sequencing.

2.3 Total RNA sequencing and quality assessment

Oulactis muscosa, *Oulactis* sp. and *Rhodactis* sp. total RNA extractions were sequenced on Illumina HiSeq 4000 – 101bp, Paired-end reads, double barcoded. Libraries were constructed using TruSeq Stranded mRNA LT sample Prep Kit. *Dofleinia* cf. *armata* samples were sequenced on Illumina HiSeq2500, 100bp Paired End reads and individual's barcoded. Libraries were constructed using PolyA enriched RNA-seq Illumina stranded sample prep. All returned libraries were quality assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

2.4 *De novo* transcriptome assembly and assessment

Raw read libraries were assembled *de novo* using Trinityrnaseq (v2.4.0) (Grabherr et al., 2011), normalised data with Trinity default settings. Sequences and Illumina adapters were trimmed using Trimmomatic (v0.36) (Bolger, Lohse, & Usadel, 2014). Transcriptome assembly quality was assessed against the number of contigs, maximum contig length and N50. The N50 was calculated for all assemblies following the Trinity method of calculation.

The percentage of reads mapping back to assembled contigs was measured using Salmon (v0.8.2) quasi-mapping method for *de novo* assembly (Patro, Duggal, Love, Irizarry, & Kingsford, 2017).

Assemblies were quantitatively assessed using Benchmarking Universal Single Copy Orthologs (BUSCO) v2.0.1 against the Eukaryote lineage database (*E*-value threshold: 10^{-5}) (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). Versions of software utilised by BUSCO were NCBI BLAST+(2.3.0) (Altschul, Gish, Miller, Myers, & Lipman, 1990), AUGUSTUS (v3.2.1) (Stanke, Steinkamp, Waack, & Morgenstern, 2004) and HMMER (v3.1b2) (Finn, Clements, & Eddy, 2011).

2.5 Additional transcriptome data

Additional transcriptomic data was sourced from publicly available databases. These included two additional species of sea anemones from the family Edwardsiidae, *Nematostella vectensis* and *Edwardsiella lineata* (Lubinski et al., 2014; Stefanik et al., 2014). Included in the analysis were the transcriptomes of two clades of the symbiotic algae *Symbiodinium* spp., clade A (CassKB8) and clade B (Mf1.05b) (Bayer et al., 2012).

2.6 Profile hidden Markov models

pHMM models were downloaded for each of the ten families from the Pfam database defensin/myotoxin-like superfamily (CL0075) (Finn et al., 2014). An additional custom profile HMM was constructed using mature peptides from full-length defensin-like sequences isolated from the tentacle transcriptomes of *Oulactis* sp. (to be published). The mature peptide sequences were aligned in ClustalΩ (Sievers et al., 2011) using Stockholm output. The hmmbuild tool in the HMMER v3.1b2 (Finn et al., 2011) suite was used to create the custom pHMM.

All transcriptome assemblies were translated into six open reading frames (ORFs). pHMM models were run against translated transcriptomes (*E*-value threshold: 10^{-10}). pHMM hits returned against contigs were examined and full-length sequences retained (start and stop codon present). The signal peptides of full-length sequences were predicted using SignalP 4.1 Server and the predicted mature peptide retained for analysis (Petersen, Brunak, von Heijne, & Nielsen, 2011).

2.7 Logo plots

Logo plots were generated using Weblogo 3 (Crooks, 2004), sequences were aligned in ClustalΩ (Sievers et al., 2011) using Stockholm output.

2.8 Alignment

Multiple sequence alignments were generated using CysBar and ClustalΩ (Shafee, Robinson, van der Weerden, & Anderson, 2016; Sievers et al., 2011). Alternative alignments were compared with Align Stat (Shafee & Cooke, 2016). Homologous cysteines were barcoded to constrain the alignment. The CysBar webserver was also used to calculate the length, charge and hydrophobicity of each sequence. For big defensins, only their defensin-like domain was included, and their N-terminal hydrophobic domain omitted.

2.9 Sequence space map

Mature defensin sequences were downloaded from Uniprot and loaded into R for analysis. Owing to the short sequence length, frequent insertions and deletions, low sequence conservation and evolutionary plasticity of cysteine-rich peptides, conventional phylogenetic methods suffer from saturation effects. We therefore also analysed the full dataset of defensin-like peptides by position-specific biophysical property sequence-space analyses (Shafee & Anderson, 2018). The variables used were R-group molecular mass (Daltons), net charge (Coulombs), hydrophobicity (Doolittle index) (Kyte & Doolittle, 1982), disorder propensity (TOP-IDP) (Campen et al., 2008), disulphide potential (binary descriptor), and Multiple Sequence Alignment (MSA) column occupancy (binary descriptor). MSA gap positions (occupancy = 0) were given the column average values for each property (other than occupancy) such that those properties did not affect dimensionality reduction. The numerical sequence space was projected by Principal Component Analysis (PCA) in [R] (Development Core Team R, 2011). The loadings of the resulting principal component axes therefore summarise the key covarying properties that most separate the sequences. Bayesian clustering was performed using Mclust (identifying optimal number, size and orientation of clusters based on goodness of fit) to identify groups of sequences with similar biophysical properties. The same process was repeated with column subsets of the MSA to confirm that clustering was not an artefact of the lack of the usually conserved cysteines and glycine.

2.10 Structural similarity dendrogram

Structural similarity was calculated as described previously (Shafee, Lay, et al., 2016). Pairwise structural similarity was calculated by the combinatorial extension method using proCKSI (http://www.procksi.net/). Hierarchical clustering was used to calculate a neighbour-joining tree based on the pairwise matrix, and presented as a dendrogram.

2.11 Visualisation

Structures were visualised for figures with Pymol. All other data were visualised with custom [R] scripts based on rgl, ggplots, igraph, and phytools (Adler, Nenadi, & Zucchini, 2003; Csárdi & Nepusz, 2006; Revell, 2012; Wickham, 2009).

3. Results

3.1 Sequencing of cnidarians and identification of *trans*-defensin sequences

During analysis of venom components of tentacle transcriptomes (to be published) for the speckled anemone (*Oulactis* sp.), numerous sequences were identified as homologous to two families in the protein PFAM Clan (CL0075). Thirteen of those sequences were most similar to the Defensin_4 family (PF07936) and one sequence to the Toxin_4 family (PF00706) (**Fig. S1**).

The defensin-like sequences found in *Oulactis* sp. were analysed for their physico-chemical properties in context with existing *trans*-defensin data (Shafee, Lay, et al., 2016). An initial sequence space analysis showed that the sequences clustered within their own region of the superfamily (**Fig. S2**). On the basis of this preliminary result, the search for defensin-like sequences was extended to newly sequenced cnidarian transcriptomes, previously published transcriptomes and cnidarian defensin sequences.

3.2 Transcriptome assemblies

Seventeen transcriptomes were assembled *de novo* from Illumina NGS data, from two orders of Cnidaria. These included one species of Corallimorpharia, *Rhodactis* sp. (identified to genus only) from the family Discosomidae, and three species of Actiniaria from the family Actiniidae: *Oulactis* sp. (undescribed species), *O. muscosa* and *Dofleinia* cf. *armata*. Various tissue samples were sequenced from each species including: tentacles (inner and outer, dependant on the species morphology), mesenterial filaments/gametes and the acrorhagi/frill for *Oulactis* species.

Overall the quality is very high for all transcriptome assemblies, with the exception of OSP10MGA (**Table S1**). Additional material was unavailable to repeat the RNA extraction for OSP10MGA. Summary statistics of the assembly, including total number of reads, number of contigs, longest contig, N50 and % of reads mapping back to contigs may be found in Supplementary **Table S1**. Assembly quality was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) which also reflected the high quality (**Fig. S3**). As the assembly for OSP10MGA contained a proportion of conserved eukaryote genes (69/404) for OSP10MGA it was kept for downstream analysis.

3.3 Defensin-like sequences gathered via profile hidden Markov models

During examination of the homologous defensin-like sequences from the *Oulactis* sp. tentacle transcriptomes, a set of highly conserved residues was identified within the BDS-like sequences (Defensin_4 family). In addition to the GxxW motif in inter-cysteine loop 2, they also began loop 3 with CP, and ended loop 4 with ICC ($CPx_{[8]}Cx_{[4]}ICC$). This subset of sequences, with the highly conserved residues, was used to build a custom pHMM (OspDef.hmm – **Fig. S4**) and run against all transcriptomes. In addition to the custom pHMM those representing each of the defensin families found in the Pfam database were run against all transcriptomes. Eleven defensin family pHMMs were run against 23 transcriptomes.

There were 210 hits returned against 21 transcriptomes for two of the eleven families of defensins and the custom pHMM (see **Fig. S5** for logo plots); Defensin_4 (42 hits), Toxin_4 (39 hits) and the custom OspDef.hmm (129 hits) (**Table S2**). Unsurprisingly, the greatest number of individual hits was for OspDef.hmm against *Oulactis* sp. transcriptomes, followed by *O. muscosa*. OspDef.hmm hits were returned against *D.* cf. *armata* and *Rhodactis* sp., although in lesser numbers.

No hits were returned for any defensin pHMMs against the reference transcriptome for the sea anemones *N. vectensis* and *E. lineata*, nor the two *Symbiodinium* sp. clades. As the Toxin_4 family pHMM is seeded on a toxin found in *N. vectensis* this was an unexpected result. In light of this, the stringency for pHMMs was reduced to an *E*-value threshold: 10^{-5} and rerun for *N. vectensis* and *E. lineata*. This less stringent setting for pHMMs still failed to yield any hits for those two reference transcriptomes.

Matched contigs were filtered for full-length sequences only (i.e. both start and stop codon present) and mature peptides retained for analysis. 32 sequences were found that matched these criteria, in addition to two partial sequences, which were retained in the analysis as they contained a portion of the pro-region peptide, a discernible cleavage site for the mature peptide and a stop codon. This resulted in a final list of 24 novel defensin-like sequences to parse into the evolution study. As *D.* cf. *armata* hits were all partial sequences (fragments) none were retained for further analysis. Novel defensin-like sequences used for physico-chemical analysis have been lodged in EMBL_ENA (European Nucleotide Archive) under project accession PRJEB27894; individual sequence accession numbers and full-length translated sequences may be found in **Table S3**.

3.4 Chemical diversity of the superfamily, and location of the cnidarian peptides

The MSA produced for cnidarian defensin-like peptides and homologues (**Fig. S6**) was used to generate a quantitative sequence space map of the *trans*-defensin superfamily. This map showed seven distinct clusters as fitted by Bayesian model-based clustering (**Fig. 1 A, B**).

Fig. 1. Sequence space PCA. (**A**) Sequences coloured by their Bayesian cluster with the main group of cnidarian peptides (blue), the big defensins (black), the α-defensins (white), the θ-defensins (red), and three groups of β-defensins (yellow, purple and grey). (**B**) Bayesian goodness-of-fit for different numbers of clusters modelled on the data. Red link indicates optimal number of clusters; error bars indicate stdev of 10 bootstrap repeats. (**C**) Sequence space PCA of the cnidarian peptides cluster only. Sequences coloured by their Bayesian subcluster with APETx-like (blue), Nv1-like (yellow), BDS-like (red), and ShI-like (white). (**D**) Bayesian goodness-of-fit for different numbers of subclusters modelled on the cnidarian peptide cluster data. Red link indicates optimal number of sub clusters; error bars indicate stdev of 10 bootstrap repeats. PC1: increasing hydrophobicity in loop 4, PC2: loss of charge and increasing disorder at loop 3, PC3: shift of hydrophobicity to loop 3 and disorder to loop 1.

The axes of the PCA-simplified sequence space summarise the biophysical properties that distinguish the main clusters (**Table S4**). Axis PC1 separates out the β and big defensins, which have a set of unified properties including hydrophobic residues earlier in loop 4, large hydrophobes at the start of loop 1, and overall bias towards residues with higher disorder propensity (**Table S4**). PC2 separates the α-defensins from the cnidarian cluster; key properties include a loss of charge and gain of disorder at start of loop 3. PC3 further separates the cnidarian cluster, especially from big defensins; key properties include a shift of disorder to earlier in loop 3 and a more hydrophobic loop 1.

The β-defensins consist of three clusters (**Fig. 1**). The cluster that contains human HBD3 (purple) is particularly separated from the others due to its distinct properties. The other two β-defensin groups have a more continuous boundary. The grey cluster contains HBD1, 2 and 136, and the majority of non-mammalian β-defensins, and the yellow cluster contains the remaining HBDs, mammalian β-defensins and the ovDLPs, crotamine, helianthamide and the 2-domain tick carboxypeptidase inhibitor. The toxic β-defensin-like proteins are located in disparate parts of the cluster, indicative of the breadth of ways to achieve toxic activity, and in agreement with their likely independent recruitment to toxic function (Chagot et al., 2005; Coronado et al., 2013; Fry et al., 2010; Roelants et al., 2010; Torres et al., 1999). The big defensins also lie in this cluster, very close to the β-defensins, indicating highly similar physico-chemical properties. Conversely, the α-defensin, θ-defensin, and cnidarian clusters are much more separated, with clear voids between them.

The cnidarian cluster of peptides (blue) has several defining properties that separate it from the wider superfamily (**Table S5**). Loop 1 is short, in the manner of α-defensins, whereas loops 3 and 4 have similar lengths to those in β-defensin. Loop 2 has markedly greater and more varied length than in other families (**Fig. 2**), and is highly flexible in solved structures

(Driscoll et al., 1989; Wilcox, Fogh, & Norton, 1993). A conserved GxxW motif is buried in solved structures and most likely important for stability. Loop 3 contains multiple small hydrophobic residues. Indeed, smaller residues are enriched at MSA columns 1, 15, 28, 43, 45, 57, and 68. Many members also share a conserved Pro-Trp pair (columns 31 and 41), which form a hydrophobic interaction in solved structures of several cnidarian homologues (PDB id: 1BDS, 1WXN, 1WQK, 1APF, 1AHL, 2H9X). Conversely, they also lack several otherwise-conserved features found in the rest of the superfamily, such as the GxC motif and the cation-anion residue pair usually present at the start of loop 2 and middle of loop 3.

Fig. 2. Inter-cysteine loop lengths for main families (three β-defensin groups from Fig. 1 combined). Bars indicate standard deviation.

This cnidarian peptide cluster can be further sub-clustered into four smaller groups: the ShI-like, BDS-like, APETx-like and Nv1-like cnidarian peptides (**Fig. 1C,D**). These subclusters have identifiable variations on the family's cysteine motif (**Table 1**) and favour slightly different sets of properties in addition to the shared features of the cnidarian cluster (**Table S6**). The BDS and Nv1 sequences also share a conserved GDxWxxR motif (**Fig. S7**).

3.5 Structural context of the cnidarian peptide cluster

We also investigated whether the cnidarian peptides are more structurally similar to each another than they are to other superfamily members with solved structures (**Fig. 3**). Hierarchical clustering of the known structures by their pairwise similarity largely agreed with clusters identified by Bayesian clustering of the sequence space (**Fig. S8**). The θ-defensin structure is too small to be reliably clustered by sequence similarity, but it is known to be most closely related to the α-defensins on the basis of its gene precursor organisation (Li et al., 2014; Shafee, Lay, et al., 2016). The α-defensins are split into two groups by the length and twist of their β-sheets (**Fig. S9**), even though the sequence space unified them in a single cluster (**Fig. S8**), which is in agreement with their shared cysteine motif.

Fig. 3. Structural similarity dendrogram of the *trans*-defensin superfamily and representative structures of the known cnidarian members. (**A**) Unrooted dendrogram based on combinatorial extension structural similarity. Main groups indicated by shape: α-defensin-like (triangles), β-defensin-like (circles), big defensin (diamond), cnidarian toxin-like (square). Vertebrate members in dark blue, invertebrate members in light blue. Antimicrobial proteins in filled shapes, toxins and inhibitors in open shapes. θ-defensin omitted due to insufficient size and unusual ligated dimer structure. Representative structures of cnidarian members: (**B**) *S. helianthus* helianthamide (4X0N) (Kem, Parten, Pennington, Price, & Dunn, 1989), (**C**) *S. helianthus* neurotoxin ShI (1SHI) (Wilcox et al., 1993), (**D**) *A. sulcata* peptide

BDS-I (1BDS) (Driscoll et al., 1989). β-strands are shown in blue, α-helices in red, and disulphide bridges in yellow.

The cnidarian peptides group into three sets by structural similarity. The helianthamide structure falls within the β-defensins, in agreement with its position in the sequence space map. This is due in part to the orientation of its α-helix, which is located in a homologous region to that of the big defensin, different from the helix found in β-defensins (**Fig. 4**). Conversely, the ShI-like and BDS-like structures group together. There is particularly close agreement for the cnidarian defensins between their structural similarity and similarity in their physico-chemical properties, resulting in neat bifurcation of the structure similarity dendrogram through that region of the sequence space map (**Fig. S9**).

Fig. 4. The α-helix (in red) that encompasses Cys_1 in the classic β-defensins (**A**) is not homologous to the α-helix that encompasses Cys_4 in (**B**) helianthamide and (**C**) the big defensins. N-terminal domain of big defensin is shown faded for clarity. PDB ids: 1FD4 (human) (Hoover et al., 2000), 4X0N (sea anemone) (Tysoe et al., 2016), 2RNG (horseshoe crab) (Kouno et al., 2008).

3.6 Taxonomic distribution of identified families

Peptides from the *trans*-defensin superfamily are distributed broadly throughout the animal kingdom, with particular expansion in the vertebrates. As yet, no homologues have been reported in the plants, fungi, or other eukaryotes. Within the phylum Cnidaria, members of the superfamily have only been detected in the class Anthozoa (sea anemones), and not the class Medusozoa (jellyfish, hydra etc.) or other cnidarians (**Fig. 5**). Most sequencing efforts have focused on the largest sea anemone family, the Actiniidae, although the helianthamide-like proteins have been found in the family Stichodactylidae.

Fig. 5. Taxonomic distribution of *trans*-defensin superfamily members. Taxonomy of phyla vertebrate lineages in dark blue, and invertebrate lineages in light blue. More detail on the family distribution of sea anemone members on the right.

4. Discussion

4.1 Taxonomic distribution and biological source

Defensin-like sequences identified in the tentacles of the sea anemone *Oulactis* sp. had the greatest similarity to the PFAM protein families of Defensin 4 and less so to Toxin 4 (Nv1). In addition to a GxxW motif, several of the sequences within the Defensins 4 family shared the additional motif ($CPx_{[8]}Cx_{[4]}ICC$), which together form a minimal hydrophobic core for those proteins. The custom pHMM based on these sequences from *Oulactis* sp. detected

further homologues in two additional species of the family Actiniidae and a species of Corallimorpharian. Sequences with this motif were restricted to the cnidarian orders Corallimorpharia and Actiniaria but not present in *N. vectensis*, *E. lineata* or the symbiotic zooxanthellae algae, *Symbiodinium* spp. (Bayer et al., 2012; Lubinski et al., 2014; Stefanik et al., 2014).

It was surprising to not detect any defensin homologues in the published reference transcriptomes for *N. vectensis* and *E. lineata*, especially as the Toxin_4 family is based on a *N. vectensis* sequence. For this reason the *E*-value threshold was raised and HMMs run again, though still without success. The lack of defensin-like sequences in these transcriptomes may be due to low venom expression at the time of tissue sampling, as venom-containing cnidae may take three to five days to be replaced subsequent to discharge (Madio, Undheim, & King, 2017; Schmidt, 1982). Defensins also occur in the mucus of a sea anemone (Sintsova et al., 2018), making it further surprising that no defensin homologues were detected in these transcriptomes. This demonstrates the necessity of sample preparation conducive to venom expression for transcriptomic-based venom evolution studies.

The precise source of the defensin-like sequences in cnidarians remains to be determined, i.e. are they located in the venom of the cnidae, the mucus or within tissue cells? Present data would indicate that they are produced by the Cnidaria themselves, rather than by symbiotic algae. Our lack of detection of *trans*-defensins in *Symbiodinium* spp. is consistent with a previous study of their AMPs that similarly did not identify any defensin-like peptides (van de Water et al., 2018). Defensin-like sequences were also detected in the tentacles of *Dofleinia* cf. *armata*, which lack any symbiotic algae. This study also confirms that cnidarian defensins do not occur solely in the tentacles; the common perception is that most toxins are located in the tentacles or in specialised fighting apparatus, such as acrorhagi. Defensin-like sequences were detected in all tissue types, external and internal, for *Rhodactis* sp. and *Oulactis* species.

4.2 Clusters separate out known functions

Within the sequence diversity of the *trans*-defensin superfamily, sequence space analysis was able to identify sets of important covarying properties. These properties segregate sequences into several clusters separated by relative voids where sequences with intermediate properties are not observed. Both the physico-chemical properties and tertiary structures of the ion-channel modulatory cnidarian peptides clearly mark them as distinct from the antimicrobial classes of the superfamily. This stems both from features specific to the cluster and the absence of features conserved in the other classes. Indeed, this cluster has some of the closest correlation of structural similarity with distribution in sequence space. They are also separated from the α-amylase inhibitory helianthamide-like proteins (Sintsova et al., 2018;

Tysoe et al., 2016). This is in contrast with the channel-binding members from vertebrates (snake, bearded lizard, and platypus), which are distributed throughout the β-defensin cluster and are much more similar to their antimicrobial homologues.

Within the cnidarian cluster, there are several discernible subgroups (ShI-like, BDS-like, APETx-like and Nv1-like) which may reflect distinct specialised functions. Sea anemone toxins such as BDS-I, BDS-II and APETx1 are known to be inactivators of Kv3.4 channels (Diochot, Loret, Bruhn, Béress, & Lazdunski, 2003; Diochot, Schweitz, Béress, & Lazdunski, 1998), although they may affect sodium as well as K_V channels (Liu, Jo, & Bean, 2012; Llewellyn & Norton, 1991). ShI and Nv1 from *Nematostella vectensis* bind to voltage-gated sodium channels, although Nv1 is more effective against insect sodium channels than mammalian (Kem et al., 1989; Moran et al., 2008). This may explain why the Nv1-like and ShI-like subgroups are separated from the BDS-like and APETx1-like subgroups in chemical space.

Several cnidarian defensins have been shown to be both ion-channel modulators and antimicrobial (e.g. crassicorin I and II from the sea anemone *Urticina crassicornis*, located in the APETx1-like subgroup) (Kim et al., 2017). Future research will be required to determine how widespread such polyfunctionality is, and whether such peptides are enriched in particular clusters in the sequence or in the interfaces between different clusters of functions. These subgroups within the cnidarian cluster may therefore be useful to guide future bioactive peptide mining, which will also help to define the sequence-function relationships and recruitment from antimicrobial to toxin function in cnidarians.

4.3 Early evolution of the superfamily

There remain a few possibilities for the early evolutionary events of the superfamily, since all the analyses presented are unrooted. One possibility is that the cnidarian members are a basal group for the superfamily. This would seem to be in line with their taxonomic distribution at the base of the Animalia. Alternatively, the big defensins may be basal, with the cnidarian peptides representing recruitment to toxic function. This possibility is in line with the multiple recruitments to toxic function of members of the superfamily already known in the vertebrates. Such repurposing of existing folds to toxic function is also common in other protein superfamilies when selection pressures are applied to evolvable protein folds (Whittington et al., 2008). The similarity of helianthamide to the big defensins in both structure and sequence properties suggests that they are probably more closely related to each other than to the other cnidarian sequences.

Additional sequencing efforts in Cnidaria (including other corals, hydroids, and scyphozoans) as well as other basal Animalia (such as Porifera) will aid in better defining the order of

evolutionary events. The present work expands our knowledge of cnidarian sequences, which have previously been lacking, and exemplifies the need to broaden venom evolutionary studies to ensure that this basal group is captured in analyses.

5. Conflict of Interest

The authors declare there are no conflicts of interest

6. Acknowledgment

This project was funded in part by ARC linkage grant LP150100621. M.L.M acknowledges an Australian Government Research Training Program Scholarship, Monash Medicinal Chemistry Faculty Scholarship and Monash University-Museum Victoria Scholarship top-up. R.S.N acknowledges fellowship support from the Australian National Health and Medical Research Council. For field work assistance in Queensland we thank Merrick Ekins, Monique Grol and Anita George.

7. References

Adler, D., Nenadi, O., & Zucchini, W. (2003). RGL : A R-library for 3D visualization with OpenGL The RGL – package. In *Proceedings of the 35th Symposium of the Interface: Computing Science and Statistics* (Vol. 35, pp. 1–11).

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arolas, J. L., Popowicz, G. M., Lorenzo, J., Sommerhoff, C. P., Huber, R., Aviles, F. X., & Holak, T. A. (2005). The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode. *Journal of Molecular Biology*, *350*(3), 489–498. https://doi.org/10.1016/j.jmb.2005.05.015

Bayer, T., Aranda, M., Sunagawa, S., Yum, L. K., Desalvo, M. K., Lindquist, E., … Medina, M. (2012). *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PloS One*, *7*(4), e35269. https://doi.org/10.1371/journal.pone.0035269

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Campen, A., Williams, R., Brown, C., Meng, J., Uversky, V., & Dunker, A. (2008). TOP-IDP-Scale: A new amino acid scale measuring propensity for intrinsic disorder. *Protein*

& Peptide Letters, *15*(9), 956–963. https://doi.org/10.2174/092986608785849164

Chagot, B., Diochot, S., Pimentel, C., Lazdunski, M., & Darbon, H. (2005). Solution structure of APETx1 from the sea anemone *Anthopleura elegantissima*: a new fold for an HERG toxin. *Proteins*, *59*(2), 380–386. https://doi.org/10.1002/prot.20425

Coronado, M. A., Gabdulkhakov, A., Georgieva, D., Sankaran, B., Murakami, M. T., Arni, R. K., & Betzel, C. (2013). Structure of the polypeptide crotamine from the Brazilian rattlesnake *Crotalus durissus terrificus*. *Acta Crystallographica Section D Biological Crystallography*, *69*(10), 1958–1964. https://doi.org/10.1107/S0907444913018003

Crooks, G. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Development Core Team R. (2011). R: A language and environment for statistical computing.

Diochot, S., Loret, E., Bruhn, T., Béress, L., & Lazdunski, M. (2003). APETx1, a new toxin from the sea anemone *Anthopleura elegantissima*, blocks voltage-gated human Ether-a-go-go-related gene potassium channels. *Molecular Pharmacology*, *64*(1), 59–69. https://doi.org/10.1124/mol.64.1.59

Diochot, S., Schweitz, H., Béress, L., & Lazdunski, M. (1998). Sea anemone peptides with a specific blocking activity against the fast inactivating potassium channel Kv3.4. *Journal of Biological Chemistry*, *273*(12), 6744–6749. https://doi.org/10.1074/jbc.273.12.6744

Driscoll, P. C., Gronenborn, A. M., Béress, L., & Clore, G. M. (1989). Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: A study using nuclear magnetic resonance and hybrid distance geometry–dynamical simulated annealing. *Biochemistry*, *28*(5), 2188–2198. https://doi.org/10.1021/bi00431a033

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., … Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(D1), D222–D230. https://doi.org/10.1093/nar/gkt1223

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl), W29–W37. https://doi.org/10.1093/nar/gkr367

Fry, B. G., Roelants, K., Winter, K., Hodgson, W. C., Griesman, L., Kwok, H. F., … Norman, J. A. (2010). Novel venom proteins produced by differential domain-expression strategies in beaded lizards and gila monsters (genus *Heloderma*). *Molecular Biology and Evolution*, *27*(2), 395–407. https://doi.org/10.1093/molbev/msp251

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Hoover, D. M., Rajashankar, K. R., Blumenthal, R., Puri, A., Oppenheim, J. J., Chertov, O., & Lubkowski, J. (2000). The structure of human β-defensin-2 shows evidence of higher order oligomerization. *Journal of Biological Chemistry*, *275*(42), 32911–32918. https://doi.org/10.1074/jbc.M006098200

Jung, S., Dingley, A. J., Augustin, R., Anton-Erxleben, F., Stanisak, M., Gelhaus, C., … Grötzinger, J. (2009). Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan hydra. *Journal of Biological Chemistry*, *284*(3), 1896–1905. https://doi.org/10.1074/jbc.M804713200

Kem, W. R., Parten, B., Pennington, M. W., Price, D. A., & Dunn, B. M. (1989). Isolation, characterization, and amino acid sequence of a polypeptide neurotoxin occurring in the sea anemone *Stichodactyla helianthus*. *Biochemistry*, *28*(8), 3483–3489. https://doi.org/10.1021/bi00434a050

Kim, C.-H., Lee, Y. J., Go, H.-J., Oh, H. Y., Lee, T. K., Park, J. B., & Park, N. G. (2017). Defensin-neurotoxin dyad in a basally branching metazoan sea anemone. *The FEBS Journal*, *284*(19), 3320–3338. https://doi.org/10.1111/febs.14194

Kouno, T., Fujitani, N., Mizuguchi, M., Osaki, T., Nishimura, S. I., Kawabata, S. I., … Kawano, K. (2008). A novel β-defensin structure: A potential strategy of big defensin for overcoming resistance by gram-positive bacteria. *Biochemistry*, *47*(40), 10611–10619. https://doi.org/10.1021/bi800957n

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Li, D., Zhang, L., Yin, H., Xu, H., Trask, J. S., Smith, D. G., … Zhu, Q. (2014). Evolution of primate α and θ defensins revealed by analysis of genomes. *Molecular Biology Reports*, *41*(6), 3859–3866. https://doi.org/10.1007/s11033-014-3253-z

Liu, P., Jo, S., & Bean, B. P. (2012). Modulation of neuronal sodium channels by the sea anemone peptide BDS-I. *Journal of Neurophysiology*, *107*(11), 3155–3167. https://doi.org/10.1152/jn.00785.2011

Llewellyn, L. E., & Norton, R. S. (1991). Binding of the sea anemone polypeptide BDS II to the voltage-gated sodium channel. *Biochemistry International*, *24*(5), 937–947.

Logashina, Y. A., Solstad, R. G., Mineev, K. S., Korolkova, Y. V, Mosharova, I. V, Dyachenko, I. A., … Andreev, Y. A. (2017). New disulfide-stabilized fold provides sea anemone peptide to exhibit both antimicrobial and TRPA1 potentiating properties. *Toxins*, *9*(5), 154. https://doi.org/10.3390/toxins9050154

Lubinski, T. J., Granger, B. R., Stefanik, D. J., Friedman, L. E., McAnulty, S. J., Schweiker, R. M., & Finnerty, J. R. (2014). A revised StellaBase enables comparative transcriptomic studies on multiple populations, life stages, and environmental conditions in the model cnidarian, *Nematostella vectensis*. *In Prep*.

Madio, B., Undheim, E. A. B., & King, G. F. (2017). Revisiting venom of the sea anemone *Stichodactyla haddoni*: Omics techniques reveal the complete toxin arsenal of a well-studied sea anemone genus. *Journal of Proteomics*, *166*, 83–92. https://doi.org/10.1016/j.jprot.2017.07.007

Moran, Y., Weinberger, H., Reitzel, A. M., Sullivan, J. C., Kahn, R., Gordon, D., … Gurevitz, M. (2008). Intron retention as a posttranscriptional regulatory mechanism of neurotoxin expression at early life stages of the Starlet anemone *Nematostella vectensis*. *Journal of Molecular Biology*, *380*(3), 437–443. https://doi.org/10.1016/j.jmb.2008.05.011

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology*, *9*(3), e1000602. https://doi.org/10.1371/journal.pbio.1000602

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Roelants, K., Fry, B. G., Norman, J. A., Clynen, E., Schoofs, L., & Bossuyt, F. (2010). Identical skin toxins by convergent molecular adaptation in frogs. *Current Biology*, *20*(2), 125–130. https://doi.org/10.1016/j.cub.2009.11.015

Schmidt, G. H. (1982). Replacement of discharged cnidae in the tentacles of *Anemonia Sulcata*. *Journal of the Marine Biological Association of the United Kingdom*, *62*(3), 685. https://doi.org/10.1017/S0025315400019834

Shafee, T. M. A., & Anderson, M. A. (2018). A quantitative map of protein sequence space for the cis-defensin superfamily. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty697

Shafee, T. M. A., & Cooke, I. (2016). AlignStat: a web-tool and R package for statistical comparison of alternative multiple sequence alignments. *BMC Bioinformatics*, *17*(1),

434. https://doi.org/10.1186/s12859-016-1300-6

Shafee, T. M. A., Lay, F. T., Hulett, M. D., & Anderson, M. A. (2016). The defensins consist of two independent, convergent protein superfamilies. *Molecular Biology and Evolution*, *33*(9), 2345–2356. https://doi.org/10.1093/molbev/msw106

Shafee, T. M. A., Lay, F. T., Phan, T. K., Anderson, M. A., & Hulett, M. D. (2017). Convergent evolution of defensin sequence, structure and function. *Cellular and Molecular Life Sciences*, *74*(4), 663–682. https://doi.org/10.1007/s00018-016-2344-5

Shafee, T. M. A., Robinson, A. J., van der Weerden, N., & Anderson, M. A. (2016). Structural homology guided alignment of cysteine rich proteins. *SpringerPlus*, *5*(1), 27. https://doi.org/10.1186/s40064-015-1609-z

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(539), 539. https://doi.org/10.1038/msb.2011.75

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V, & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sintsova, O., Gladkikh, I., Chausova, V., Monastyrnaya, M., Anastyuk, S., Chernikov, O., … Kozlovskaya, E. (2018). Peptide fingerprinting of the sea anemone *Heteractis magnifica* mucus revealed neurotoxins, Kunitz-type proteinase inhibitors and a new β-defensin α-amylase inhibitor. *Journal of Proteomics*, *173*, 12–21. https://doi.org/10.1016/j.jprot.2017.11.019

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*(Web Server), W309–W312. https://doi.org/10.1093/nar/gkh379

Stefanik, D. J., Lubinski, T. J., Granger, B. R., Byrd, A. L., Reitzel, A. M., DeFilippo, L., … Finnerty, J. R. (2014). Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, *Edwardsiella lineata*, a parasitic cnidarian. *BMC Genomics*, *15*(1), 71. https://doi.org/10.1186/1471-2164-15-71

Torres, A. M., Wang, X., Fletcher, J. I., Alewood, D., Alewood, P. F., Smith, R., … Kuchel, P. W. (1999). Solution structure of a defensin-like peptide from platypus venom. *Biochemical Journal*, *341*(3), 785–794. https://doi.org/10.1042/bj3410785

Tysoe, C., Williams, L. K., Keyzers, R., Nguyen, N. T., Tarling, C., Wicki, J., … Withers, S. G. (2016). Potent human α-Amylase inhibition by the β-Defensin-like protein Helianthamide. *ACS Central Science*, *2*(3), 154–161. https://doi.org/10.1021/acscentsci.5b00399

van de Water, J. A. J. M., Chaib De Mares, M., Dixon, G. B., Raina, J.-B., Willis, B. L., Bourne, D. G., & van Oppen, M. J. H. (2018). Antimicrobial and stress responses to increased temperature and bacterial pathogen challenge in the holobiont of a reef-building coral. *Molecular Ecology*, *27*(4), 1065–1080. https://doi.org/10.1111/mec.14489

Vidal-Dupiol, J., Ladrière, O., Destoumieux-Garzón, D., Sautière, P.-E., Meistertzheim, A.-L., Tambutté, E., … Mitta, G. (2011). Innate immune responses of a scleractinian coral to vibriosis. *Journal of Biological Chemistry*, *286*(25), 22688–22698. https://doi.org/10.1074/jbc.M110.216358

Whittington, C. M., Papenfuss, A. T., Bansal, P., Torres, A. M., Wong, E. S. W., Deakin, J. E., … Belov, K. (2008). Defensins and the convergent evolution of platypus and reptile venom genes. *Genome Research*, *18*(6), 986–994. https://doi.org/10.1101/gr.7149808

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Wilcox, G. R., Fogh, R. H., & Norton, R. S. (1993). Refined structure in solution of the sea anemone neurotoxin ShI. *The Journal of Biological Chemistry*, *268*(33), 24707–24719.

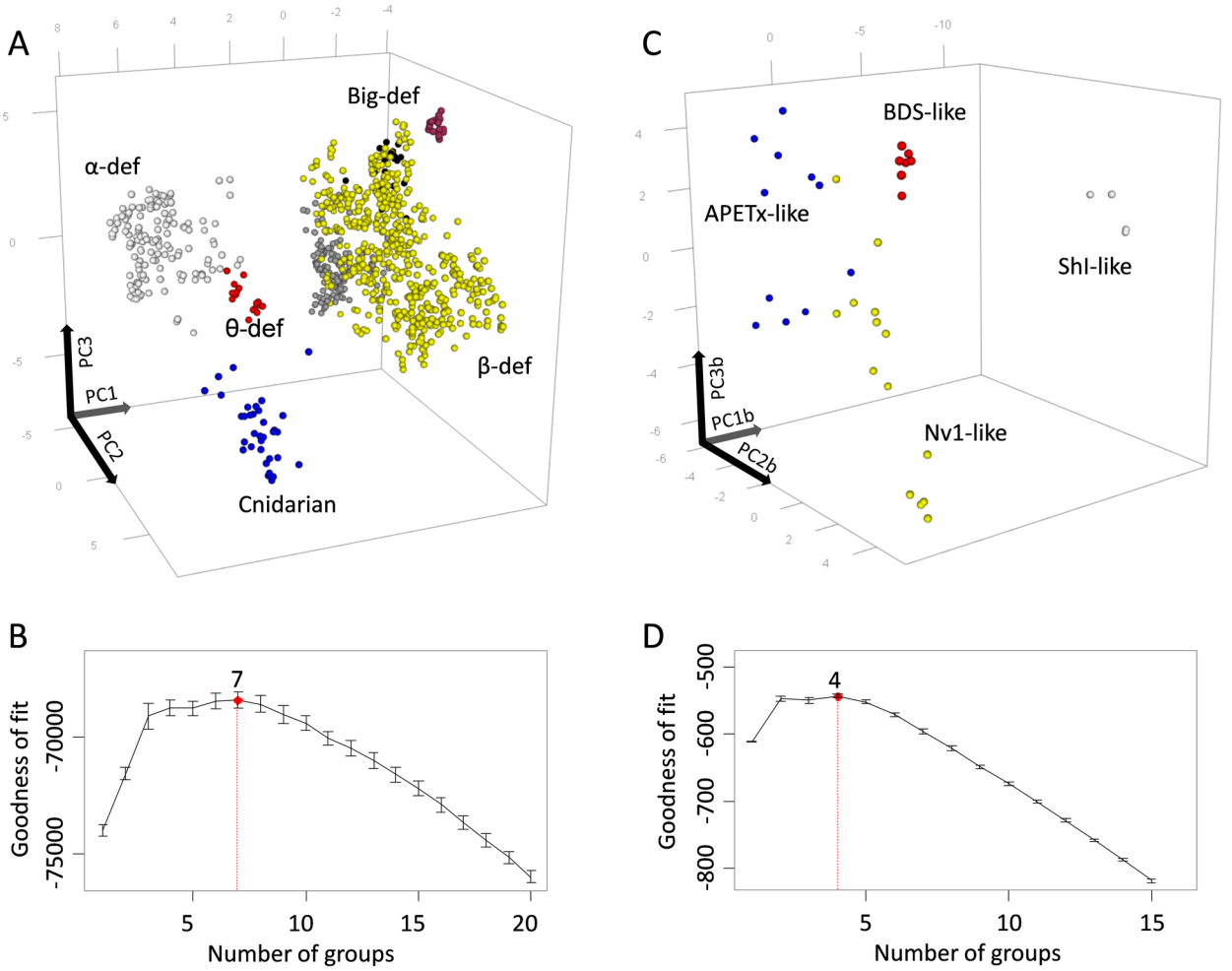
Zhu, S., & Gao, B. (2013). Evolutionary origin of β-defensins. *Developmental & Comparative Immunology*, *39*(1–2), 79–84. https://doi.org/10.1016/j.dci.2012.02.011

Zielezinski, A., Vinga, S., Almeida, J., & Karlowski, W. M. (2017). Alignment-free sequence comparison: benefits, applications, and tools. *Genome Biology*, *18*(1), 186. https://doi.org/10.1186/s13059-017-1319-7

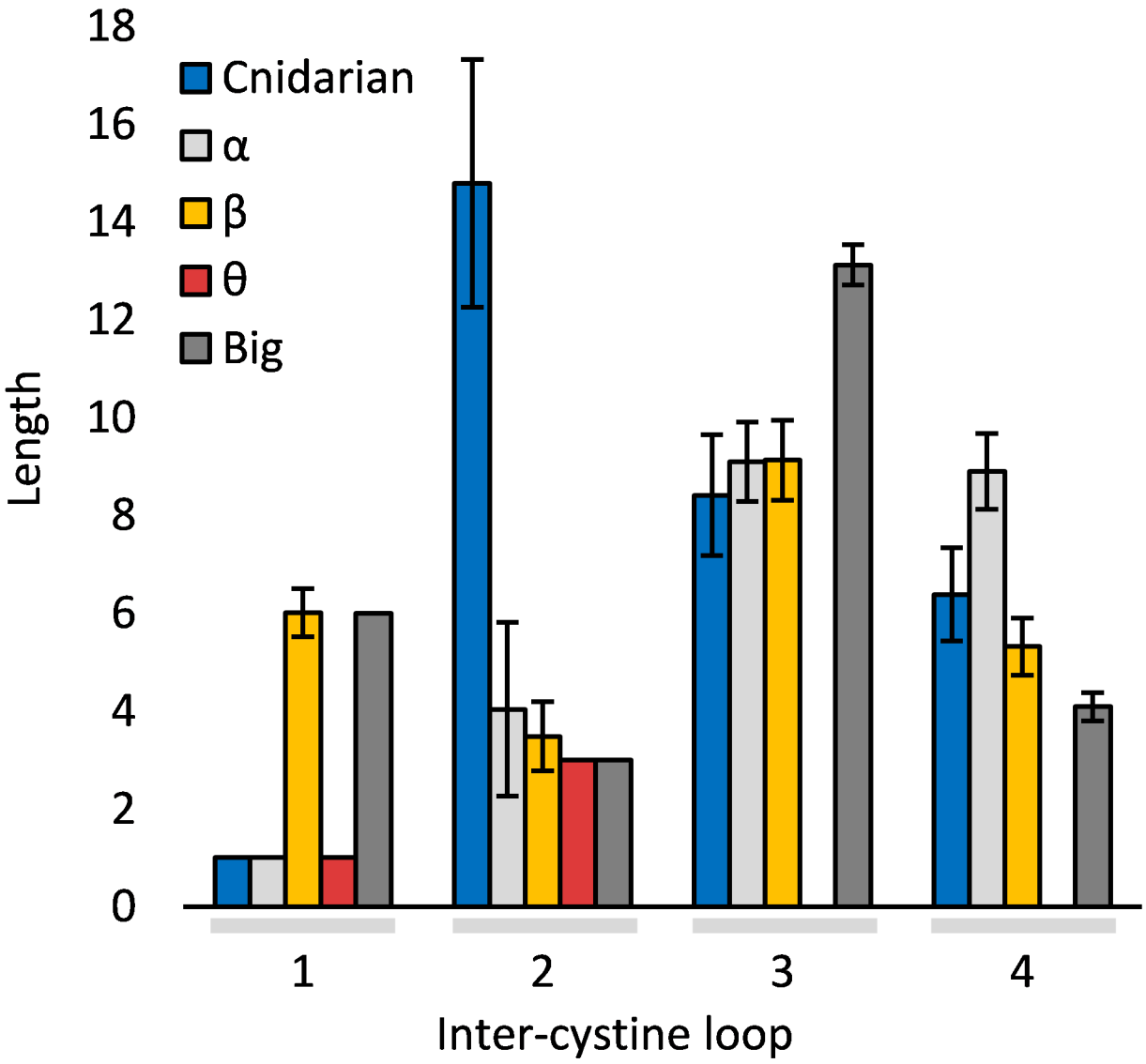
Table I. Cysteine motifs for cnidarian *trans*-defensins.

Cnidarian peptide subgroup	Cysteine motif [1]
APETx-like [2]	$X_{[2-3]}CXCX_{[12-17]}CX_{[7--9]}CX_6CCX_{[0-5]}$
Nv1-like [2]	$X_{[2-3]}CXCX_{[13-15]}CX_{[6--10]}CX_{[5-6]}CCX_{[0-5]}$
BDS-like [2]	$X_3CXCX_{[15-17]}CX_9CX_6CCX_{[0-5]}$
ShI-like [3]	$X_3CXCX_{20}CX_6CX_{[7-9]}CC_{[2-3]}$
Helianthamide-like [3]	$X_5CX_9CX_3CX_{12}CX_4CC$

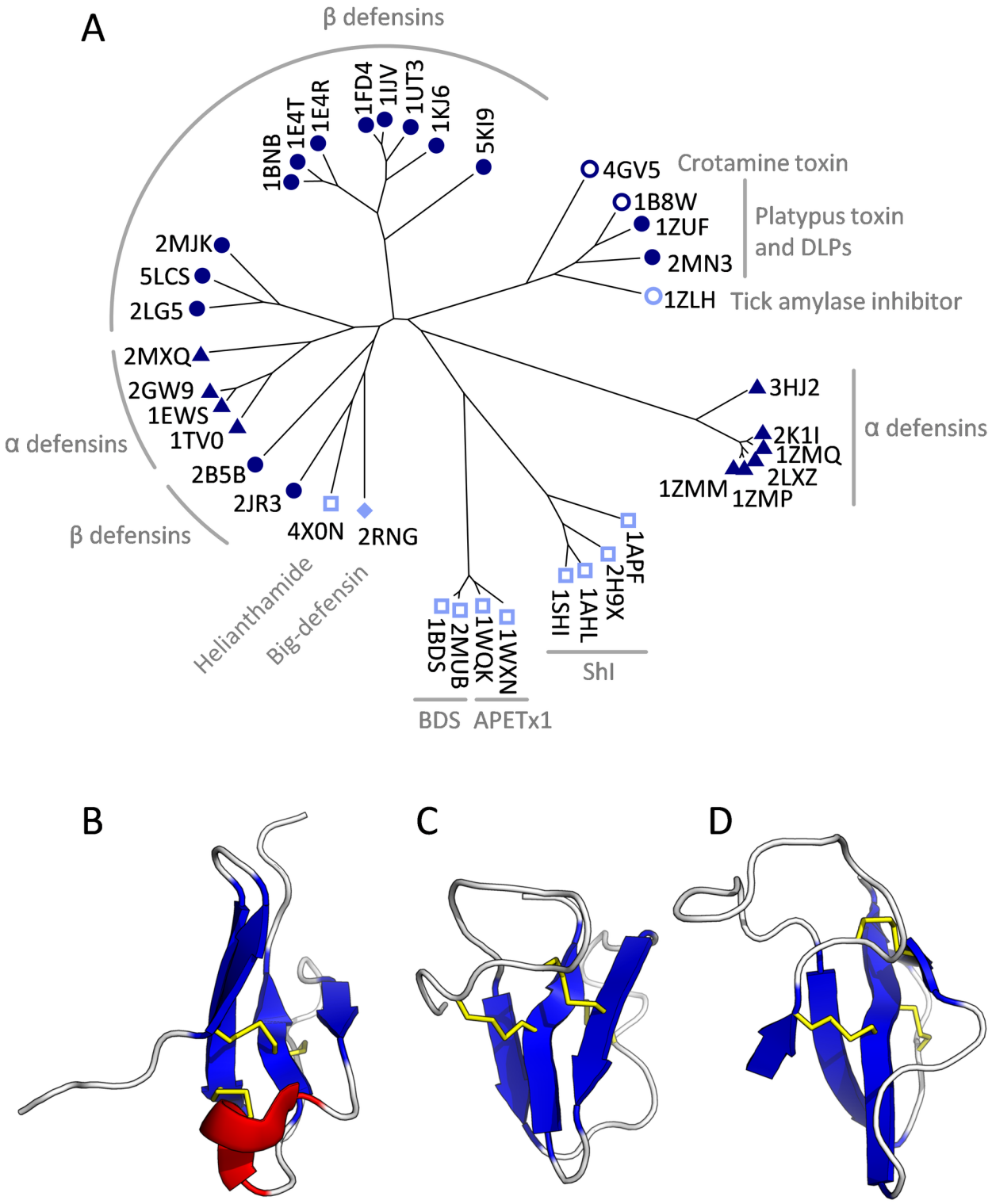
[1] Incorporating 95% of occurring variation). [2] The first four groups are subgroups of the cnidarian peptide cluster. [3] The helianthamide-like proteins cluster with the β-defensins.



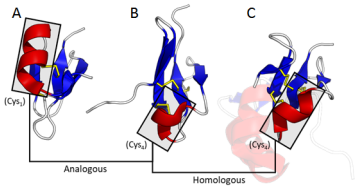
PROT_25679_Figure 1.tif



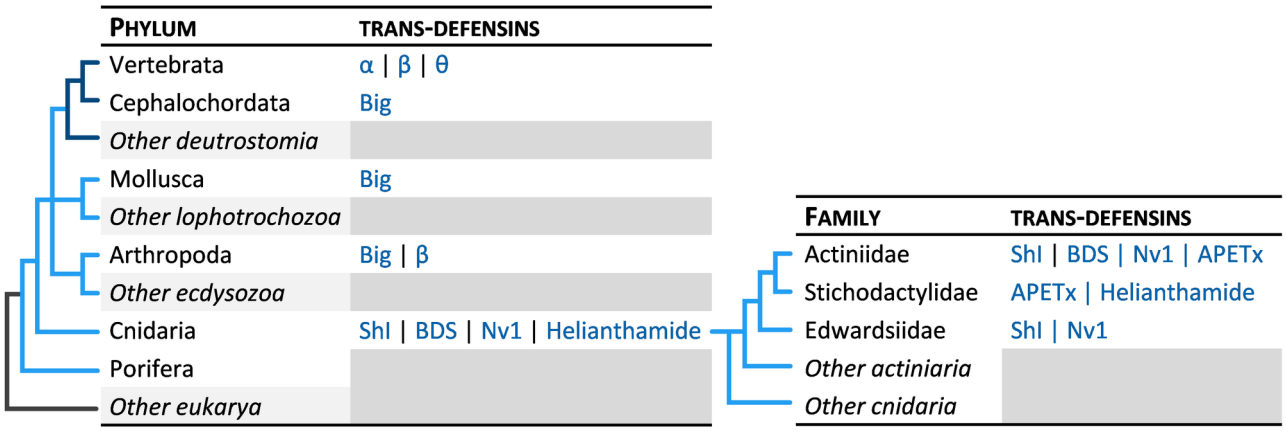
PROT_25679_Figure 2.tif



PROT_25679_Figure 3.tif



PROT_25679_Figure 4.tif



PROT_25679_Figure 5.tif

Table I. Cysteine motifs for cnidarian *trans*-defensins.

Cnidarian peptide subgroup	Cysteine motif [1]
APETx-like [2]	$X_{[2-3]}CXCX_{[12-17]}CX_{[7--9]}CX_6CCX_{[0-5]}$
Nv1-like [2]	$X_{[2-3]}CXCX_{[13-15]}CX_{[6--10]}CX_{[5-6]}CCX_{[0-5]}$
BDS-like [2]	$X_3CXCX_{[15-17]}CX_9CX_6CCX_{[0-5]}$
ShI-like [3]	$X_3CXCX_{20}CX_6CX_{[7-9]}CC_{[2-3]}$
Helianthamide-like [3]	$X_5CX_9CX_3CX_{12}CX_4CC$

[1] Incorporating 95% of occurring variation). [2] The first four groups are subgroups of the cnidarian peptide cluster. [3] The helianthamide-like proteins cluster with the β-defensins.