The neuropeptide transcriptome of a model echinoderm, the sea urchin *Strongylocentrotus purpuratus*.

Matthew L. Rowe and Maurice R. Elphick

Queen Mary University of London,
School of Biological & Chemical Sciences,
Mile End Road,
London, E1 4NS, UK

Correspondence to: Prof. M.R. Elphick,
Queen Mary University of London,
School of Biological & Chemical Sciences,
Mile End Road,
London, E1 4NS, UK

Tel: 44 207 882 5290
Fax: 44 208 983 0973
E-mail: M.R.Elphick@qmul.ac.uk

Abstract

Neuronal secretion of peptide signaling molecules (neuropeptides) is an evolutionarily ancient feature of nervous systems. Here we report the identification of twenty cDNAs encoding putative neuropeptide precursors in the sea urchin *Strongylocentrotus purpuratus* (Phylum Echinodermata), providing new insights on the evolution and diversity of neuropeptides. Identification of a gonadotropin-releasing hormone-like peptide precursor (SpGnRHP) is consistent with the widespread phylogenetic distribution of GnRH-type neuropeptides in the bilateria. A protein (SpTRHLP) comprising multiple copies of peptides that share structural similarity with thyrotropin-releasing hormone (TRH) is the first TRH-like precursor to be identified in an invertebrate. SpCTLP is the first calcitonin-like peptide with two N-terminally located cysteine residues to be found in a non-chordate species. Discovery of two proteins (SpPPLNP1, SpPPLNP2) comprising homologs of molluscan pedal peptides and arthropod orcokinins indicates the existence of a bilaterian family of pedal peptide/orcokinin-type neuropeptides. Other proteins identified contain peptides that do not share apparent sequence similarity with known neuropeptides. These include Spnp5, which comprises multiple copies of C-terminally amidated peptides that have an N-terminal Ala-Asn motif (AN peptides), and Spnp9, Spnp10 and Spnp12, which contain putative neuropeptides with a C-terminal Phe-amide, Ser-amide or Pro-amide, respectively. Several proteins (Spnp 11, 14, 15, 16, 17, 18, 19 and 20) contain putative neuropeptides with multiple cysteine residues (2, 6 or 8), which may mediate formation of intramolecular or intermolecular disulphide bridges. Looking ahead, the identification of these neuropeptide precursors in *Strongylocentrotus purpuratus* has provided a strong basis for a comprehensive analysis of neuropeptide function in this model echinoderm species.

Keywords: Neuropeptide; echinoderm; sea urchin; *Strongylocentrotus purpuratus*; evolution

1. Introduction

Neuronal secretion of peptide signaling molecules (neuropeptides) is a fundamental and evolutionarily ancient feature of nervous systems. Unlike "classical" neurotransmitters (e.g. acetylcholine, dopamine), which are synthesized by enzymes, neuropeptides are cleaved from precursor proteins and therefore mutation-induced changes in the amino acid sequences of neuropeptides can occur over time [42, 75]. Accordingly, it has been proposed that changes in the sequences and/or the expression of neuropeptide genes may be important in the evolution of behavior, with neuropeptide genes acting as "volume knobs" that shape adaptive changes in animal behavior over evolutionary time [5]. Consistent with this notion, neuropeptides act as mediators and/or regulators of a wide range of behaviors, including locomotor activity, feeding, reproduction and learning [35, 61].

Neuropeptides were first discovered on account of their effects as neurohormones on physiological phenomena such as blood pressure or the contractile activity of visceral organs and only later was it found that these molecules also act within the central nervous system to regulate whole-animal behavior. For example, the neuropeptides vasopressin and oxytocin were discovered as pituitary neurohormones that cause an increase in blood pressure and uterine contraction, respectively [19]. Subsequently it was found that vasopressin-releasing and oxytocin-releasing neurons also project to many regions of the central nervous system [15] and both vasopressin and oxytocin are now known to be key players in neural mechanisms of social behavior [54, 91].

Other technical strategies for neuropeptide discovery that have been important include the use of antibodies to known neuropeptides to enable identification of structurally related neuropeptides in the same species or in other species [17, 18] and the use of mass spectroscopic and/or sequencing techniques to identify putative bioactive neuropeptides in

87 extracts of neural tissue [29]. More recently, however, it is has been the use of genome
88 sequencing and/or sequencing of neural cDNA libraries that has transformed neuropeptide
89 discovery, with comprehensive genome-wide analyses of putative neuropeptide precursor
90 genes being accomplished in several animal species. For example, when the first animal
91 genome sequences were obtained for the model organisms *Caenorhabditis elegans* [77] and
92 *Drosophila melanogaster* [1], detailed surveys of candidate neuropeptide precursor genes
93 were reported [34, 48, 81]. These initial overviews of neuropeptide diversity in
94 *Caenorhabditis elegans* and *Drosophila melanogaster* provided the foundations for
95 subsequent more detailed studies that have identified the receptors that mediate the effects of
96 neuropeptides in these animals and have provided new insights on the
97 physiological/behavioral roles of neuropeptides [33, 47, 53].
98 During the last decade or so genome sequencing technology has been applied to an
99 increasingly wide range of animal species and genome-wide surveys of neuropeptide diversity
100 have been reported for species belonging to several animal phyla, including annelids [14, 84],
101 molluscs [83] and cnidarians [4]. Thus, a picture of the diversity of neuropeptides that occur
102 throughout the animal kingdom and the relationships between these neuropeptides is
103 beginning to emerge. The picture is still far from complete but we can see on the horizon the
104 potential for reconstructing the evolutionary history of neuropeptide signaling systems based
105 upon detailed comparative analysis of the complements and characteristics of neuropeptides
106 in extant species.
107 Animals that are important for investigation of the phylogenetic distribution and
108 evolution of neuropeptides are the deuterostomian invertebrates because they provide a link
109 between vertebrates (also deuterostomes) and the protostomian invertebrates, which include
110 phyla such as arthropods (e.g. *Drosophila*), nematodes (e.g. *C. elegans*), molluscs and
111 annelids. Genome sequences have been obtained for urochordate and cephalochordate species

112 [16, 66], which are of particular interest because these invertebrate chordates are the closest
113 extant relatives of the vertebrates. Furthermore, surveys of neuropeptide diversity have been
114 reported for the urochordate *Ciona intestinalis* [32, 45, 71, 73]. Aside from these invertebrate
115 chordate sub-phyla, three deuterostomian invertebrate phyla are currently recognised: the
116 echinoderms, the hemichordates and the xenacoelomorphs [8, 9, 64]. A genome-sequencing
117 project is on going for a hemichordate species [28] but a genome-sequencing project has been
118 completed for an echinoderm species, the sea urchin *Strongylocentrotus purpuratus* [74].
119 Thus, analysis of neuropeptide genes in this species and in other echinoderm species has the
120 potential to provide important insights on the evolution of neuropeptides in the
121 deuterostomian branch of the animal kingdom.
122 Pioneering studies on neuropeptides in echinoderms detected a peptide in extracts of
123 starfish nerve cords that triggers gamete maturation and release – "gamete-shedding
124 substance" or "gonad-stimulating substance" (GSS) [12, 43] However, the molecular identity
125 of GSS remained unknown for fifty years until it was identified as a relaxin-like peptide in
126 2009 [59]. Another neuropeptide that triggers gamete maturation and release has been
127 identified in the sea cucumber *Apostichopus japonicus* as NGIWYamide, which is structurally
128 unrelated to relaxin-like GSS in starfish [44]. Interestingly, NGIWYamide was discovered
129 previously as one of a number of neuropeptides that were identified in *Apostichopus* on
130 account of their effects on the contractility of *in vitro* preparations of body wall muscle and/or
131 intestine from this species [39, 40]. Furthermore, the protein precursors of NGIWYamide and
132 other myoactive neuropeptides in *Apostichopus* have recently been identified by analysis of
133 transcriptome sequence data [21]. However, the first neuropeptides to be identified in an
134 echinoderm, the SALMFamides S1 and S2, were isolated on account of their cross-reactivity
135 with antibodies to the molluscan neuropeptide pQDPFLRFamide [23, 24]. S1 and S2 were
136 both purified from extracts of radial nerve cords from the starfish species *Asterias rubens* and

137 *Asterias forbesi* and subsequent studies have revealed that S1, S2 and SALMFamide
138 neuropeptides identified in other echinoderms act as muscle relaxants [22, 57].
139 Sequencing of the *Strongylocentrotus purpuratus* genome [74] has provided the first
140 opportunity for a comprehensive analysis of neuropeptide diversity in an echinoderm species.
141 Identification of thirty-eight genes encoding putative neuropeptide receptors or peptide
142 hormone receptors [10] indicates that the diversity of neuropeptide signaling pathways in this
143 echinoderm species is comparable to findings from species belonging to other invertebrate
144 phyla [34]. However, analysis of the *Strongylocentrotus purpuratus* genome sequence data
145 using search strategies such as the Basic Local Alignment Search Tool (tBLASTn; [2]) only
146 revealed a handful of putative neuropeptide genes. These included a gene encoding seven
147 putative SALMFamide neuropeptides [26], a gene encoding a vasopressin/oxytocin-type
148 neuropeptide ("echinotocin") [25], a gene encoding two copies of a peptide (NGFFFamide)
149 related to the sea cucumber neuropeptide NGIWYamide [25], three homologs of glycoprotein
150 hormones (SpGPH1; SpGPH2; SpGPH3 [10]), and two genes encoding homologs of the two
151 subunits that form the insect neurohormone bursicon [10]. This paucity of putative
152 neuropeptide genes identified based on BLAST analysis of genome sequence data suggests
153 that many other neuropeptide genes in *Strongylocentrotus purpuratus* remain to be discovered
154 and furthermore that other strategies are needed to identify these genes in the genome.
155 One strategy that has been successfully employed to identify putative neuropeptides in
156 *Strongylocentrotus purpuratus* is the use of mass spectrometry and genomic database
157 searching to identify and sequence neuropeptides [58]. Here we present a complementary
158 strategy, namely the analysis of expressed sequence tag (EST) data obtained from a
159 *Strongylocentrotus purpuratus* nerve cord cDNA library. We recently demonstrated the utility
160 of this approach with the identification of a second SALMFamide precursor gene in
161 *Strongylocentrotus purpuratus* (Spnp1; [68]) and here we have extended the use of this

approach with the identification of nineteen other putative neuropeptide precursor genes (Spnp2 – Spnp20). The data presented here provide novel insights on the evolution of neuropeptide signaling systems as well as providing a basis for studies in which the expression and physiological/behavioral roles of neuropeptides are investigated in a model echinoderm species.

2. Materials and methods

The sequences of 2026 expressed sequence tags (ESTs) derived from a *Strongylocentrotus purpuratus* radial nerve cDNA library were downloaded from the National Center for Biotechnology Information (NCBI) EST database (dbEST). These included 1027 3' reads (GI:109401590 - 109402616) and 999 5' reads (GI:109402617-109403615), all approximately 1000 nucleotides in length.

To identify transcripts encoding putative neuropeptide precursors the 5' EST dataset was first selected for analysis because 5' untranslated regions (UTRs) are typically shorter than 3' UTRs [63] and therefore 5' ESTs usually contain more coding sequence than 3' ESTs. A recent analysis of RNAseq data has confirmed this for *Strongylocentrotus purpuratus*, with an average 5' UTR length of 269 base pairs and an average 3' UTR length of 1799 base pairs [80]. The 5' ESTs were analysed by submission as queries against the GenBank protein database using BLASTx and ESTs encoding proteins that were clearly identifiable as homologs of known proteins that are not neuropeptide precursors were discarded from further analysis.

A N-terminal signal peptide is required for targeting of neuropeptide precursors to the lumen of the endoplasmic reticulum as the first step towards the regulated secretory pathway [75]. Therefore, employing the online signal peptide prediction tool SignalP (http://www.cbs.dtu.dk/services/SignalP/; [6]), absence of an N-terminal signal peptide sequence was used as a second criterion for further elimination of ESTs encoding proteins that are not neuropeptide precursors.

Neuropeptide precursors are typically quite small proteins (e.g. 50-500 residues) and therefore the length of proteins encoded by ESTs was also used as a criterion for assessment of potential neuropeptide precursors. Furthermore, more detailed analysis of ESTs encoding

197 candidate neuropeptide precursors involved inspection of their primary amino acid sequences
198 to identify their potential neuropeptide products by searching for the presence of sequences
199 bounded by potential dibasic (KR, RR, KK, RK) as well as monobasic (R) endopeptidase
200 cleavage sites [70, 82]. The presence of a glycine residue preceding the putative C-terminal
201 cleavage site was noted as a potential substrate for C-terminal amidation [20]. Likewise, the
202 presence of a N-terminal glutamine residue (Q) was noted as a potential substrate for post-
203 translational conversion to a pyroglutamate (pQ) residue [27]. The presence of cysteine
204 residues was also noted, recognising the potential for the formation of intramolecular or
205 intermolecular disulphide bridges.
206 A subset of ESTs encoding twenty putative neuropeptide precursor proteins was
207 identified. Full length radial nerve cDNA sequences were obtained, where possible, by
208 combining 5' EST sequences with 3' EST sequence data, which was obtained by submission
209 of the predicted neuropeptide precursors as BLAST queries against dbEST
210 (http://www.ncbi.nlm.nih.gov/dbEST/). This BLAST search also enabled identification of
211 those putative neuropeptide precursors that are also expressed in other adult tissues or in other
212 development stages in *Strongylocentrotus purpuratus*.
213 Radial nerve cDNA sequences encoding putative neuropeptide precursors were also
214 subject to further analysis to obtain definitive sequences by comparison with genomic
215 sequence data using the BLAST facility on SpBase (http://sugp.caltech.edu/SpBase/; [11]). In
216 particular, the aim here was to correct any EST sequencing errors and also to determine the
217 exon-intron structure of genes encoding the putative neuropeptide precursors by identification
218 of 5' (gt) and 3' (ag) consensus sites for intron splicing. SpBase was also used to determine if
219 putative neuropeptide precursors were predicted by the gene prediction tool GLEAN3, which
220 was used for gene annotation during the annotation phase of the sea urchin genome project
221 (http://www.hgsc.bcm.tmc.edu/projects/seaurchin; [74]). Likewise, the NCBI sea urchin

genome resource was used to determine if putative neuropeptide precursors were predicted by the gene prediction tool Gnomon (http://www.ncbi.nlm.nih.gov/projects/genome/guide/sea_urchin/).

Having obtained definitive sequences for the radial nerve cDNAs encoding putative neuropeptide precursors based on combined EST and genomic sequence data, additional 5' and 3' sequence data was obtained by submission of the radial nerve cDNAs as queries in BLAST searches of RNAseq data obtained from a variety of sea urchin tissues, including radial nerve, and available for BLAST analysis on the SpBase website (http://sugp.caltech.edu/SpBase/rnaseq/; [80]). Thus, the transcript sequences encoding putative neuropeptide precursors that are shown in the supplementary figures of this paper include a core sequence based on the original radial nerve EST sequence data (not underlined) with additional 5' and 3' sequences obtained from RNAseq data underlined.

3. Results and Discussion

We report here the identification of twenty putative neuropeptide precursors in the sea urchin *Strongylocentrotus purpuratus*. The strategy employed involved analysis of the sequences of 2026 ESTs derived from a radial nerve cDNA library, complementing a previous study that used mass spectrometric analysis of radial nerve extracts with reference to the *Strongylocentrotus purpuratus* genome sequence [58]. Analysis of the radial nerve EST dataset revealed cDNAs encoding two neuropeptide precursors that were originally discovered by analysis of genomic sequence data using BLAST: the F-type SALMFamide precursor [26] and the NGFFFamide precursor [25]. These findings demonstrated that known neuropeptide precursors are represented amongst the collection of 2026 radial nerve ESTs analysed, providing an important indication that more detailed scrutiny of the EST dataset might reveal additional neuropeptide precursors. The first novel neuropeptide precursor identified by analysis of the EST dataset was the L-type SALMFamide precursor reported previously [68], which we have designated as "*Strongylocentrotus purpuratus* neuropeptide precursor 1" or Spnp1 (Fig. 1 and Fig. S1). Here we report the discovery of a further nineteen putative neuropeptide precursors, which we have designated Spnp2 – Spnp20. The sequences of these precursor proteins are shown in Fig. 1 and Fig. 2, whilst in supplementary figures S1 – S20 the cDNA sequences are included together with their translated protein products. A detailed description and discussion of Spnp2 – Spnp20 is presented below.

3.1. Spnp2 (SpGNRHP): precursor of a gonadotropin-releasing hormone-type peptide

Spnp2 is a 131-residue protein comprising a predicted 30-residue N-terminal signal peptide followed by a gonadotropin-releasing hormone (GnRH)-like neuropeptide (SpGnRH) with the predicted sequence pyroGlu-Val-His-His-Arg-Phe-Ser-Gly-Trp-Arg-Pro-Gly-NH_2

(Fig. 1). The N-terminal pyroGlu residue and the C-terminal amide group are predicted based on the occurrence of these post-translational modifications in GnRH-type peptides identified in other species and the existence of this GnRH-type peptide in *Strongylocentrotus purpuratus* has been reported previously [67].

The protein sequence of Spnp2 (SpGnRH precursor or SpGnRHP) was initially identified by analysis of the sequences of the radial nerve cDNAs RNSP-1M3 (5': EC439573.1, GI:109403596, 3': EC438289.1, GI:109402312), RNSP-1D2 (5': EC439527.1, GI:109403550, 3': EC428144.1, GI:109402167), RNSP-1N3 (5': EC439440.1, GI:109403463, 3': EC438444.1, GI:109402467), RNSP-9I7 (5': EC439133.1, GI:109403156) and RNSP-1G17 (5': EC439418.1, GI:109403441, 3': EC438392.1, GI:109402415). However, a cDNA encoding Spnp2 is also represented in a larval cDNA library (MPMGp691D2380, 5': CD294893.1, GI:34745970, 3': EC437745.1, GI:109401768). Furthermore, Spnp2 was predicted from automated analysis of genomic sequence data by gene prediction tools (Gnomon - GI:72011734; GLEAN3_19680) and assigned the gene ID number SPU_019680 [74]. The Spnp2 cDNA sequence shown in Fig. S2 is a consensus sequence derived from genomic, cDNA/EST and RNAseq (WHL22.157157.0) sequence data.

GnRH was originally discovered in mammals on account of its stimulatory effect on the release from the anterior pituitary of the gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH) [69]. Subsequently, GnRH-type peptides have been identified in other vertebrates [46] and in invertebrates, including molluscs [41], annelids [84] and urochordates [31, 79]. Furthermore, adipokinetic hormone (AKH) – type peptides are homologs of GnRH found in arthropods and nematodes [49, 50]. Thus, the discovery of a gene encoding a GnRH-like peptide in *Strongylocentrotus purpuratus* was to be expected because the GnRH neuropeptide family has a widespread phylogenetic distribution in the animal kingdom. However, SpGnRH is the first member of the GnRH neuropeptide family to

be identified in an echinoderm. The existence of SpGnRH in *Strongylocentrotus purpuratus* has recently been reported independently as part of a broad analysis of the evolution of the GnRH neuropeptide family [67]. Roch et al. (2011) compared the sequence of SpGnRH with the sequences of identified or putative GnRH-type peptides in a wide range of animal phyla, highlighting similarities with GnRH-type peptides in vertebrates, invertebrate chordates and other invertebrate phyla.

The protein-coding region of the SpGnRH gene comprises three exons with the signal peptide and SpGnRH peptide encoded by the first exon (Fig. S2), a feature that is shared with the human GnRH gene and GnRH genes in other vertebrates. The presence of single copy of the GnRH peptide is a feature of all known GnRH-type precursors, although the vertebrate precursors include a C-terminal gonadotropin-associated peptide (GAP) region containing a 52 residue protein of unknown function. The SpGnRH precursor also contains an 85 residue C-terminal sequence, which spans all three exons, but which lacks sequence homology to vertebrate GAP.

It is of interest to consider the potential physiological roles of SpGnRH in sea urchins. A recurring theme for GnRH-type neuropeptides throughout the animal kingdom is a role in regulation of reproductive processes. For example, GnRH-immunoreactivity is present in a nerve plexus that innervates the gonads and gonoducts of the urochordate *Ciona intestinalis* [52] and GnRH-type peptides stimulate gamete release in *Ciona* [76]. Furthermore, in the nematode *C. elegans* RNAi-mediated knockdown of genes encoding a GnRH/AKH-type peptide or a GnRH-type receptor delays egg-laying [50]. Investigation of the physiological roles of SpGnRH in sea urchins would be facilitated by identification of the receptor(s) that mediate effects of this peptide. Candidate receptors are proteins that share sequence similarity with the G-protein coupled GnRH-type receptors (GnRHRs) that have been identified in vertebrates and other invertebrates [67]. Relevant in this regard are genes encoding three

GnRHR-like proteins in *Strongylocentrotus purpuratus* - SpGnRHR1 (GI:185134933; SPU_001536), SpGnRHR2 (GI:185134985; SPU_001537); and SpGNRHR3 (GI:185134947; SPU_001531). Characterisation of the ligand-binding properties of these proteins and analysis of their tissue/organ expression profiles in *Strongylocentrotus purpuratus* may facilitate investigation of the physiological roles of SpGnRH in sea urchins.

3.2. Spnp3 (SpTRHLP): precursor of a thyrotropin-releasing hormone-like peptide

Spnp3 is a 316-residue precursor protein comprising a predicted 15-residue N-terminal signal peptide and nineteen putative neuropeptides bounded by monobasic or dibasic cleavage sites (Fig. 1). These include ten copies of the sequence QYPGG, four copies of the sequence QWPGG and single copies of the sequences QFPAG, QFPGG, QFVGGELIPSPEL, QWPEV and QFVGGEALEQESNIN. The presence of a N-terminal glutamine (Q) residue and a C-terminal glycine (G) residue in the majority of these sequences are indicative of post-translational modifications giving rise to a N-terminal pyroglutamate residue (pQ) and C-terminal amide group. For example, the most abundant of the putative neuropeptide sequences (QYPGG) would give rise to mature peptides with the structure pGlu-Tyr-Pro-Gly-NH_2. This peptide is noteworthy because it shares structural similarity with human thyrotropin-releasing hormone (TRH, pGlu-His-Pro-NH_2). Therefore, we refer to Spnp3 as SpTRH-like precursor (SpTRHLP) and we refer to the most abundant of its putative constituent peptides (pGlu-Tyr-Pro-Gly-NH_2) as SpTRH.

The protein sequence of Spnp3 was initially identified by analysis of the sequences of the radial nerve cDNA RNSP-9P21 (5': EC438846.1, GI:109402869, 3': EC437745.1, GI: 109401768). However, a cDNA encoding Spnp3 is also represented in a primary mesenchyme cell cDNA library (PMCSPR2-184H11, 5': DN579827.1 GI:61138866). Furthermore, Spnp3 was predicted from automated analysis of genomic sequence data by

gene prediction tools (Gnomon - GI:185134999; GLEAN3_08352) and assigned the gene ID number SPU_008352 [74]. The Spnp3 cDNA sequence shown in Fig. S3 is a consensus sequence derived from genomic, cDNA/EST and also RNAseq (WHL22.3018.0) sequence data.

Thyrotropin-releasing hormone (TRH; pGlu-His-Pro-NH_2) is a hypothalamic hormone that stimulates release of thyroid stimulating hormone (TSH) and prolactin from the anterior pituitary. In humans and most mammals, six copies of TRH are derived from the TRH precursor protein, whereas in non-mammalian vertebrates eight copies of TRH appears to be the norm [85]. Although TRH-like immunoreactivity has been detected in a crustacean species [37], to the best of our knowledge TRH-like peptides have not been identified in any invertebrate species.

Spnp3 encodes a 316-residue protein comprising ten copies of the sequence QYPGG, which shares structural similarity with TRH. Thus, with post-translational conversion of the N-terminal glutamine residue to pyroglutamate and use of the C-terminal glycine as a substrate for amidation, a peptide would be formed (pGlu-Tyr-Pro-Gly-NH_2; SpTRH) that has a pGlu-X-Pro motif, as found in TRH. The peptide pGlu-Tyr-Pro-Gly-NH_2 also shares C-terminal sequence similarity with SpGnRH (pGlu-Val-His-His-Arg-Phe-Ser-Gly-Trp-Arg-Pro-Gly-NH_2) and other GnRH-type peptides. However, the sea urchin peptide and is more similar to TRH than GnRH both in terms of its length (four residues) and the existence of multiple copies of the peptide in its precursor. Nevertheless, the similarity with GnRH is intriguing and it may perhaps indicate that TRH-type peptides originated from a GnRH-type peptide. SpTRHLP is the first putative precursor of a TRH-like peptide to be discovered in an invertebrate species, indicating that the origins of the TRH-type peptides may date back at least as far as the common ancestor of deuterostomes.

Investigation of the physiological roles of SpTRH in sea urchin would be facilitated by identification of the receptor that this putative peptide binds to. In mammals TRH exerts effects by binding to a G-protein coupled TRH receptor and a gene encoding a protein that is closely related to mammalian TRH receptors has been identified in *Strongylocentrotus purpuratus* (SPU_010167; [10]). Therefore, SPU_010167 is a candidate mediator of the effects of the SpTRH peptide in sea urchins. Interestingly, whilst TRH-like peptides have thus far only been found in vertebrates, TRH receptor-like proteins have been identified in protostomian invertebrates. For example, the *Drosophila* gene CG2114 encodes an ortholog of vertebrate TRH receptors and the endogenous ligands for this receptor are FMRFamide-type neuropeptides [56]. Thus, TRH-type receptors date back to the common ancestor of the bilateria and it appears that amidated short peptides have evolved as endogenous ligands for these receptors in different branches of the animal kingdom. Therefore, it will be interesting to determine if the putative amidated tetrapeptide, SpTRH (pGlu-Tyr-Pro-Gly-NH_2), is indeed the endogenous ligand for the TRH receptor-like protein (SPU_010167) in sea urchins.

3.3. Spnp4 (SpCTLPP): precursor of a calcitonin-like peptide

Spnp4 is a 110-residue protein comprising a predicted 21-residue N-terminal signal peptide and, bounded by dibasic cleavage sites, a 38-residue peptide that shares structural similarity with calcitonin and calcitonin-like peptides (Fig. 1). The C-terminal residue of the 38-residue peptide is a glycine residue, which is a potential substrate for C-terminal amidation. Thus, Spnp4 is predicted to give rise to a 37-residue peptide with the sequence SKGCGSFSGCMQMEVAKNRVAALLRNSNAHLFGLNGP-NH_2, which we refer to as *Strongylocentrotus purpuratus* calcitonin-like peptide (SpCTLP).

The protein sequence of Spnp4 was initially identified by analysis of the sequences of the radial nerve cDNAs RNSP-9A2 (5': EC438671.1, GI:109402694, 3': EC437655.1,

383 GI:109401678), RNSP-5C1 (5': EC439006.1, GI:109403029, 3': EC438242.1, GI:109402265),
384 RNSP-5H24 (5': EC439062.1, GI:109403085; 3': EC437612.1, GI:109401635), RNSP-9I12
385 (5': EC438743.1, GI:109402766, 3': EC437839.1, GI:109401862), RNSP-5P22 (5':
386 EC439097.1, GI:109403120, 3': EC437635.1, GI:109401658), RNSP-9M18 (5': EC438785.1,
387 GI:109402808, 3': EC437878.1, GI:109401901) and RNSP-9K20 (5': EC438638.1,
388 GI:109402661, 3': EC437888.1, GI:109401911). However, cDNAs encoding Spnp4 are also
389 represented in a larval cDNA library (MPMGp691H2032, 5': CD309678.1, GI:34754727,
390 MPMGp691H16126, CD307674.1, GI:34752723). Furthermore, Spnp4 was predicted from
391 automated analysis of genomic sequence data by gene prediction tool Gnomon
392 (GI:115767208). The Spnp4 cDNA sequence shown in Fig. S4 is a consensus sequence
393 derived from genomic and cDNA/EST sequence data.
394 In mammals calcitonin is released from parafollicular cells of the thyroid gland and
395 inhibits Ca^{2+} absorption by the intestines and osteoclast activity in bones [87]. Calcitonin is
396 encoded by a gene that also encodes calcitonin-gene related peptide (CGRP), with alternative
397 splicing of transcripts giving rise to either prepro-calcitonin (exons 1, 2, 3 and 4) or prepro-
398 CGRP (exons 1, 2, 3, 5 and 6) [3]. By way of comparison, the Spnp4 has five exons, the fifth
399 of which encodes SpCTLP peptide sequence (Figure S4).
400 A characteristic that SpCTLP shares with both calcitonin and CGRP is the presence of
401 two cysteine residues in the N-terminal region of the peptide (residues 4 and 10 in SpCTLP)
402 (Fig. 3). In calcitonin and CGRP these cysteine residues form a disulphide bridge and it seems
403 likely, therefore, that this is also a feature of SpCTLP. At the C-terminus of SpCTLP is a
404 putative Pro-amide motif and in this respect SpCTLP is more like calcitonin than CGRP (see
405 Fig. 3).
406 Calcitonin/CGRP-like peptides have been identified throughout the vertebrates [62]
407 but relatively little is known about the occurrence and characteristics of calcitonin/CGRP-type

408 peptides in invertebrates. A key finding was the discovery that a diuretic hormone (DH31)
409 identified in the cockroach *Diploptera punctate* is structurally related to calcitonin [30]
410 providing important molecular evidence that calcitonin/CGRP-type peptides may have a
411 widespread phylogenetic distribution in the animal kingdom. However, DH31-type peptides
412 identified in *Diploptera punctate* and in other insects do not have the two cysteines that are a
413 feature of the N-terminal region of calcitonin/CGRP-type peptides in vertebrates. Recently a
414 calcitonin-like peptide (Ci-CT) was identified in the sea-squirt *Ciona intestinalis* (Phylum
415 Chordata) and this peptide does have two cysteine residues in its N-terminal region, which
416 indicated that this feature may be a unique characteristic of calcitonin/CGRP-type peptides in
417 chordates [72]. It is of interest, therefore, that the calcitonin-like peptide identified here in the
418 sea urchin *Strongylocentrotus purpuratus* (SpCTLP) also has two cysteine residues, which are
419 located at positions 4 and 10 (Fig. 3). This suggests that this feature of calcitonin/CGRP-type
420 peptides in vertebrates can in fact be traced back beyond the chordates to the common
421 ancestor of extant deuterostomes.
422 As SpCTLP is the first calcitonin-like peptide to be discovered in an echinoderm it
423 will be interesting to investigate its physiological roles. Opportunities to do this would be
424 facilitated by identification of its receptor. In mammals, calcitonin exerts effects by binding to
425 a G-protein coupled receptor that belongs to secretin-type family of receptors. A homolog of
426 mammalian calcitonin/CGRP-type receptors is present in *Strongylocentrotus purpuratus*
427 (SPU_018314; Burke et al. 2006) and therefore this is a likely candidate as the receptor for
428 SpCTLP.
429
430 *3.4. Spnp5 (SpANPP): precursor of a family of peptides with a N-terminal Ala-Asn motif – the*
431 *AN peptides*

Spnp5 is a 441-residue protein that comprises a 27-residue signal peptide and thirteen copies of putative neuropeptides that are structurally related, all having an N-terminal dipeptide sequence Ala-Asn (AN) (Fig. 1). Therefore, we have designated these as *Strongylocentrotus purpuratus* AN peptides or SpANPs and we refer to Spnp5 as *Strongylocentrotus purpuratus* AN peptide precursor (SpANPP). Spnp5 contains one copy of the sequence ANYFRGRGRKPG (SpANP1), eight copies of the sequence ANMFRSRLRGKG (SpANP2), two copies of the sequence ANMFRSRLRGNG (SpANP3), one copy of the sequence ANYFRGRGRRPG (SpANP4) and one copy of the sequence ANFRARQRPKLGK (SpANP5). It is noteworthy that all of these AN peptide sequences except ANP5 have a C-terminal glycine residue, which is a potential substrate for C-terminal amidation. Other structural characteristics are shared amongst some but not all of the putative neuropeptides.

The protein sequence of Spnp5 (SpANPP) was initially identified by analysis of the sequences of the radial nerve cDNAs RNSP-5B16 (5': EC438945.1, GI:109402968, 3': EC438118.1, GI:109402141), RNSP-5K1 (5': EC438975.1, 109402998, 3': EC438249.1, GI:109402272), RNSP-9G19 (5': EC438680.1, GI:109402703, 3': EC437567.1, GI:109401590), RNSP-9M16 (5': EC438692.1, GI: 109402715, 3': EC437817.1, GI:109401840), RNSP-5O19 (5': EC439324.1, GI:109403347, 3': EC438561.1, GI:109402584) and RNSP-5G14 (5': EC439381.1, GI:109403404, 3': EC438026.1, GI: 109402049). Interestingly, however, a cDNA encoding Spnp5 is also represented amongst cDNAs from bacterially activated coelomocytes (CK829173.1, GI:50873844). Automated analysis of genomic sequence data with gene prediction tools produced conflicting data with respect to Spnp5. The GLEAN3 method predicted that the 441 residues of Spnp5 form the C-terminal region of a much larger protein comprising 2208 residues (GLEAN3_18666), which was named Sp-Zcchc11 on account of its zinc finger motifs and a CCHC domain and was

assigned the gene ID number SPU_018666 [74]. On the other hand the Gnomon gene prediction method predicts a 441-residue protein that is identical to Spnp5 (SpANPP). Importantly, both EST and RNAseq data confirm the existence of transcripts that encode the 441-residue Spnp5 (SpANPP) protein and therefore we can conclude that the Gnomon gene prediction was correct and the GLEAN3 gene prediction was incorrect. Thus, the Spnp5 cDNA sequence shown in Fig. S5 is a consensus sequence derived from genomic, cDNA/EST and also RNAseq (WHL22.164432.1) sequence data.

Mass spectroscopic analysis has confirmed that three of the peptides predicted to be derived from SpANPP, SpANP1, SpANP2 and SpANP3, are present in nerve cords from *Strongylocentrotus purpuratus* and are C-terminally amidated [58]. Furthermore, we have independently confirmed the presence of SpANP2 in extracts of tests from *Strongylocentrotus purpuratus* (M.L. Rowe, R.D. Burke and M.R. Elphick, unpublished data). We have not identified any striking sequence similarities that AN peptides share with neuropeptides identified in other phyla. Thus, there are no comparative perspectives on potential physiological roles of these neuropeptides in sea urchins. Furthermore, preliminary pharmacological studies testing synthetic SpANP2 for myoactivity in sea urchins did not reveal effects on the contractile activity of tube foot or oesophagus preparations (M.L. Rowe and M.R. Elphick, unpublished data). Therefore, further studies are now required to investigate the physiological roles of AN peptides in sea urchins.

3.5. Spnp6 (SpPPLNP1) and Spnp7 (SpPPLNP2): precursors of peptides related to molluscan pedal peptides and arthropod orcokinins.

Both Spnp6 and Spnp7 contain putative neuropeptides that share sequence similarity with pedal peptide (PLDSVYGTHGMSGFA), a neuropeptide originally isolated from the mollusc *Aplysia californica* [51]. Therefore we refer to Spnp6 as *Strongylocentrotus*

purpuratus pedal peptide-like neuropeptide precursor 1 (SpPPLNP1) and we refer to Spnp7 as *Strongylocentrotus purpuratus* pedal peptide-like neuropeptide precursor 2 (SpPPLNP2)

Spnp6 (SpPPLNP1) is a 510-residue protein (Fig. 1) comprising a 29-residue N-terminal signal peptide and 21 copies of pedal peptide-like neuropeptides: SpPPLN1a (RFLTGALEPLSSGFI; 1 copy), SpPPLN1b (GFNTGAMEPLGSGFI; 2 copies), SpPPLN1c (GFNSGAMEPLGAGFF; 8 copies), SpPPLN1d (GFNSGAMEPLGSGFI; 5 copies), SpPPLN1e (GFNNGAMEPLGSGFI; 1 copy), SpPPLN1f (DFNTGAMEPLGSGFI; 1 copy), SpPPLN1g (GFHAGAMEPLSSGFIDG; 1 copy), SpPPLN1h (GFYNGAMEPLSAGFHQG; 1 copy) and SpPPLN1i (GFHNGAMEPLKSGFLKD; 1 copy).

The protein sequence of Spnp6 (SpPPLNP1) was initially identified by analysis of the sequences of the radial nerve cDNA RNSP-9E2 (5': EC438675.1, GI:109402698, 3': EC437660.1, GI: 109401683). However, cDNAs encoding Spnp6 are also represented in blastula cDNA libraries (MPI_537_46L9, 5': CD332062.1, GI:34798584, 3': CD324334.1, GI: 34796395; yda51d10, 5': CX558302.1, GI:57585331; yda83f08, 5': CX554074.1, GI:57581103), a primary mesenchyme cell cDNA library (PMCSPR2-101N6, 5': DN788099.1, GI:62376892, 3': DN564330.1, GI: 61123369), a gastrula cDNA library (MPI_536_18H7, 5': CD339652.1, GI:34806178) and a lantern cDNA library (LSP-2M22, 5': EC435043.1, GI:109399066, 3': EC430111.1, GI: 109394134). Furthermore, Spnp6 was predicted from automated analysis of genomic sequence data by gene prediction tools (Gnomon - GI:72008820; GLEAN3_03108) and assigned the gene ID number SPU_003108 [74]. The Spnp6 cDNA sequence shown in Fig. S6 is a consensus sequence derived from genomic, cDNA/EST and also RNAseq (WHL22.633184) sequence data.

Spnp7 (SpPPLNP2) is a 204-residue protein comprising a putative 19-residue N-terminal signal peptide and ten putative neuropeptides: SpPPLN2a (FGSMNMEPLVSGFY), SpPPLN2b (FGSGLDSMQSGFY), SpPPLN2c (NFGSGLNMEPMQSGFY), SpPPLN2d

(NFGGSMEPMQSGFY), SpPPLN2e (FGGAMEPMSSGFY), SpPPLN2f (FGSGSLEPMSSGFY; 2 copies), SpPPLN2g (NFGGSLEPMQSGFY), SpPPLN2h (FGGANEPMRSGFF) and SpPPLN2i (NFGGSLDAMQSGFY).

The protein sequence of Spnp7 was initially identified by analysis of the sequences of the radial nerve cDNA RNSP-5B10 (5': EC439068.1, GI:109403091, 3': EC438075.1, GI:109402098). However, a larval cDNA/EST encoding Spnp7 (MPMGp691F1380, 5': CD294941.1, GI:34746018, 3': CD309924.1, GI: 34754973) has also been deposited in the GenBank database. Furthermore, Spnp7 was also predicted from automated analysis of genomic sequence data using gene prediction tools (Gnomon - GI:390352582; GLEAN3_24381) and was assigned the gene ID number SPU_024381 [74]. Thus, the Spnp7 cDNA sequence shown in Fig. S7 is a consensus sequence derived from genomic, cDNA/EST and also RNAseq (WHL22.656375.0) sequence data.

Pedal peptide was originally isolated from the mollusc *Aplysia californica* and was named pedal peptide because it is predominantly synthesised in pedal ganglia of the *Aplysia* central nervous system [51]. Subsequently, the sequence of a precursor protein (pedal peptide 1 precursor) from which pedal peptide is derived has been determined [60], revealing that it contains 17 copies of the peptide originally isolated by Lloyd and Connolly [51] as well as two other structurally related peptides. Furthermore, in *Aplysia* there are three other precursors containing peptides related to pedal peptide and these are known as pedal peptide 2 precursor, pedal peptide 3 precursor and pedal peptide 4 precursor [60]. In Figure 4, the sequences of representative peptides derived from SpPPLNP1 (PPLN1d) and from SpPPLNP2 (PPLN2h) are aligned with the prototypical *Aplysia* pedal peptide and representative peptides derived from pedal peptide 2 precursor and pedal peptide 3 precursor. Both PPLN1d and PPLN2h share a C-terminal SGFX (where X is a hydrophobic residue) motif with pedal peptide but otherwise the level of sequence identity is quite low. However,

532 the sea urchin and *Aplysia* peptides have similar characteristics with respect to the number of
533 residues and distribution of hydrophobic and hydrophilic residues. Furthermore, like
534 SpPPLNP1 and SpPPLNP2 (see Fig. 1), the *Aplysia* pedal peptide-type precursors comprise
535 many copies of the constituent peptides [60]. Importantly, when SpPPLNP1 is submitted as a
536 BLAST query against the GenBank protein database it is the *Aplysia* pedal peptide 2
537 precursor that is the next best hit after SpPPLNP2. Collectively, these findings suggest that
538 SpPPLNP1 and SpPPLNP2 share a common evolutionary ancestry with the *Aplysia* pedal
539 peptide-type precursors.

540 Discovery of SpPPLNP1 and SpPPLNP2 is of particular interest because these are the
541 first pedal peptide-type neuropeptide precursors to be discovered in a deuterostomian
542 invertebrate. Thus, the existence of pedal peptide-type neuropeptide precursors in a
543 protostomian invertebrate (the mollusc *Aplysia californica*) and a deuterostomian invertebrate
544 (the echinoderm *Strongylocentrotus purpuratus*) suggests that the origins of pedal peptide-
545 type neuropeptides may trace back at least as far as the common ancestor of bilaterian
546 animals. Therefore, the existence of pedal peptide-type neuropeptide precursors in other
547 protostomian and deuterostomian animal phyla would be expected. Consistent with this
548 notion, genes encoding pedal peptide-like precursor proteins have been identified in the
549 annelid species *Capitella teleta*, *Helobdella robusta* and *Platynereis dumerilii* [14, 84] and a
550 representative pedal peptide-type neuropeptide from *Platynereis dumerilii* is included in the
551 alignment in Fig. 4. Surprisingly, however, there are no reports in the literature of pedal
552 peptide-type precursors in ecdysozoan protostomes such as arthropods and nematodes, which
553 could of course reflect loss of pedal peptide-type genes in the ecdysozoan lineage. However,
554 because our discovery of SpPPLNP1 and SpPPLNP2 has revealed that pedal peptide-type
555 precursors are not restricted to lophotrochozoan phyla (e.g. molluscs and annelids), we have
556 investigated the occurrence of pedal peptide-type precursors in nematodes and arthropods.

557 Interestingly, we have identified pedal peptide-type precursors in the nematode
558 *Caenorhabditis elegans* (NLP14; GI:392926792 and NLP15; GI:7498042). Furthermore,
559 submission of these *Caenorhabditis* pedal peptide-type precursors as BLAST queries against
560 the GenBank protein database reveals that they share significant similarity with molluscan
561 pedal peptide precursors and also with precursors of orcokinin-type neuropeptides in several
562 arthropod species. This suggests that orcokinins and pedal peptide-type neuropeptides may be
563 members of a bilaterian family of homologous neuropeptides. Accordingly, the sequences of
564 orcokinin peptides from the crustacean *Procambrus clarkii* and the insect *Nasonia vitripennis*
565 are also included in the alignment shown in Figure 4.
566 Discovery of SpPPLNP1 and SpPPLNP2 has provided a basis for investigation of the
567 physiological roles of pedal peptide-type neuropeptides in the sea urchin *Strongylocentrotus*
568 *purpuratus*. Moreover, because our findings indicate that pedal peptide/orcokinin-type
569 neuropeptides may occur throughout the bilateria, it will be of particular interest to compare
570 the functions of these peptides in protostomian and deuterostomian invertebrates.
571 Progress towards functional studies on peptides derived from SpPPLNP1 and
572 SpPPLNP2 has been facilitated by mass spectroscopic analysis of *Strongylocentrotus*
573 *purpuratus* [58]. Thus, mass spectrometry has confirmed the presence in nerve extracts of
574 eight of the peptides (SpPPLN1a – SpPPLN1h) predicted to be derived from SpPPLNP1 and
575 has also revealed that the C-terminal glycine residue of SpPPLN1g and SpPPLN1h is a
576 substrate for amidation - GFHAGAMEPLSSGFIDamide and
577 GFYNGAMEPLSAGFHQamide, respectively. Interestingly, a peptide corresponding to a C-
578 terminally truncated form of SpPPLN1i that lacks the last two residues of the predicted
579 peptide (i.e. GFHNGAMEPLKSGFL as opposed to GFHNGAMEPLKSGFLKD) was
580 detected in nerve extracts, which may indicate an unusual utilisation of lysine (K) as a
581 monobasic cleavage site. Mass spectrometry has also confirmed the presence in nerve extracts

582 of six of the nine peptides predicted to be derived from SpPPLNP2 (SpPPLN2a, SpPPLN2b,
583 SpPPLN2c, SpPPLN2d, SpPPLN2f and SpPPLN2h) [58].
584 Using HPLC-MS we have independently confirmed the presence of SpPPLN1c
585 (GFNSGAMEPLGAGFF) in extracts of tests from *Strongylocentrotus purpuratus* (M.L.
586 Rowe, R.D. Burke and M.R. Elphick, unpublished data). We have also tested synthetic
587 SpPPLN1c for myoactivity on sea urchin tube foot or oesophagus preparations, but no effects
588 were observed (M.L. Rowe and M.R. Elphick, unpublished data). Therefore, as with AN
589 peptides, further studies are now required to investigate the physiological roles of PPLN1-
590 and PPLN2-type neuropeptides in sea urchins.
591
592 *3.6. Spnp8*
593 Spnp8 is a 85-residue protein comprising a predicted 22-residue N-terminal signal
594 peptide followed by a 63-residue sequence (residues 23-85) that contains a putative dibasic
595 cleavage site (KR) at residues 55 and 56 (Fig. 1). The C-terminal region of the protein
596 (residues 57-85) contains six acidic residues (D or E), indicating that this part of the protein
597 functions as an acidic spacer peptide. On this basis we propose that it is the 32-residue
598 polypeptide formed by residues 23-54 that may be a secreted bioactive neuropeptide.
599 However, residues 23-54 also include a potential monobasic cleavage site (R), so there
600 remains the possibility that the polypeptide formed by residues 23-54 is cleaved into two
601 smaller bioactive neuropeptides.
602 The protein sequence of Spnp8 was initially identified by analysis of the sequences of
603 the 27 radial nerve cDNAs: RNSP-1H1 (5': EC439462.1; GI:109403485, 3: EC438486.1,
604 GI:109402509), RNSP-9M8 (5': EC438717.1, GI:109402740, 3': EC437865.1,
605 GI:109401888), RNSP-1B10 (5': EC438982.1, GI:109403005, 3': EC438229.1,
606 GI:109402252), RNSP-1G3 (5': EC439570.1, GI:109403593, 3': EC438336.1,

GI:109402359), RNSP-1E5 (5': EC439589.1, GI:109403612; 3': EC438296.1, GI:109402319), RNSP-1K10 (5': EC439419.1, GI:109403442, 3': EC438388.1, GI:109402411), RNSP-9H4 (5' EC438985.1, GI:109403008, 3': EC437792.1, GI:109401815), RNSP-9D1 (5': EC438727.1, GI:109402750, 3': EC437872.1, GI:109401895), RNSP-9D6 (5': EC438965.1, GI:109402988, 3': EC437772.1, GI:109401795), RNSP-1L3 (5': EC439437.1, GI:109403460, 3': EC438443.1, GI:109402466), RNSP-1K1 (5': EC439577.1, GI:109403600, 3': EC438315.1, GI:109402338), RNSP-9J15 (5': EC438756.1, GI:109402779, 3': EC437782.1, GI:109401805), RNSP-5E14 (5': EC439380.1, GI:109403403, 3': EC438958.1, GI:109402981), RNSP-9J21 (5': EC438839.1, GI:109402862, 3': EC437742.1, GI:109401765), RNSP-9A15 (5': EC438948.1, GI:109402971, 3': EC437598.1, GI:109401621), RNSP-1D1 (5': EC439459.1, GI:109403482, 3': EC438485.1, GI:109402508), RNSP-1D12 (5': EC438880.1, GI:109402903, 3': EC438199.1, GI:109402222), RNSP-5G20 (5': EC439365.1, GI:109403388, 3': EC438529.1, GI:109402552), RNSP-5K20 (5': EC439155.1, GI:109403178, 3': EC437663.1, GI:109401686), RNSP-5I6 (5': EC439250.1, GI:109403273, 3': EC437904.1, GI:109401927), RNSP-5L9 (5': EC438910.1, GI:109402933, 3': EC437976.1, GI:109401999), RNSP-5C12 (5': EC439201.1, GI:109403224, 3': EC437986.1, GI:109402009), RNSP-5I22 (5': EC439245.1, GI:109403268, 3': EC437910.1, GI:109401933), RNSP-1I3 (5': EC439571, GI:109403594), RNSP-9G12 (5': EC438741.1, GI:109402764, 3':EC437837.1, GI:109401860), RNSP-9P11(5': EC438781.1, GI:109402804, 3':EC437795.1, GI:109401818), RNSP-5G17 (5': EC439347, GI:109403370). This large number of radial nerve cDNAs encoding Spnp8 suggests that it is expressed at a high level in the adult nervous system. However, cDNAs encoding Spnp8 are also represented amongst cDNAs from larvae (MPMGp691F1913; 5': CD295824, GI:34746901) and primary mesenchyme cells (PMCSPR2-127I19; 5': DN790471.1, GI:62380538, 3': DN563688.1, GI:61122727).

Spnp8 was predicted from genome sequence data by the gene prediction tool Gnomon (XP_001175942.1, GI:115764725). It was not predicted, however, by the GLEAN3 tool that was used for genome annotation [74] and therefore it has as yet not been assigned a gene ID number. The Spnp8 cDNA sequence shown in Fig. S8 is a consensus sequence derived from genomic and cDNA/EST sequence data.

The putative 32-residue neuropeptide derived from Spnp8 does not share any apparent sequence similarity with neuropeptides or peptide hormones identified in other phyla.

3.7. Spnp9

Spnp9 is a 97-residue protein comprising a predicted 18-residue N-terminal signal peptide followed by a 79-residue sequence (residues 19-97) that contains a putative dibasic cleavage site (KR) at residues 42 and 43 (Fig. 1). The C-terminal region of the protein (residues 44-97) contains nine acidic residues (D or E), indicating that this part of the protein may function as an acidic spacer peptide. On this basis we propose that it is the 23-residue polypeptide formed by residues 19-41 that may be a secreted bioactive neuropeptide. However, residues 19-41 also include two potential monobasic cleavage sites (R), so there remains the possibility that the polypeptide formed by residues 19-41 is cleaved into two or three smaller bioactive neuropeptides. The C-terminal residue of the polypeptide sequence formed by residues 19-41 is glycine, which may be a substrate for amidation.

The protein sequence of Spnp9 was initially identified by analysis of the sequences of the radial nerve cDNAs RNSP-5J5 (3': EC437977.1, GI:109402000), RNSP-9A13 (5': EC439000.1, GI:109403023, 3': EC437636.1, GI:109401659), RNSP-9O3 (5': EC438941.1, GI:109402964, 3': EC437709.1, GI:109401732). However, cDNAs encoding Spnp9 are also represented amongst cDNAs from larvae (MPMGp691E2327, 5': CD305936.1, GI:34750985), blastulae (yda60d12, 5': CX559052.1, GI:57586081;

yde01d12, 5': CX698100.1, GI:57960911; ydd37h11, 5': CX691973.1, GI:57954046; ydc90e11, 5': CX694141.1, GI:57956476; yda10h10, 5': CX079346.1, GI:56593336; yda48h06, 5': CX199608.1, GI:56847032; ydc58c11, 5': CX681794.1, GI:57942445; yde84e05, 5': CX692690.1, GI:57954849 and primary mesenchyme cells (91222952_F24_086_PC_0025_A1_MR_C12, 5': BG780665.1, GI:14151678).

Spnp9 was not predicted from automated analysis of genomic sequence data by the gene prediction tools Gnomon and GLEAN3 and therefore Spnp9 has not been assigned a gene ID number. The Spnp9 cDNA sequence shown in Fig. S9 is a consensus sequence derived from genomic and cDNA/EST sequence data.

If the 23-residue putative neuropeptide derived from Spnp9 is amidated, then the C-terminal region of the peptide (HGMPFamide) shares sequence similarity with members of the SALMFamide neuropeptide family (e.g. AYQTGLPFamide, an L-type SALMFamide neuropeptide in the starfish *Marthasterias glacialis* [92]). This relatively low level of sequence similarity may of course reflect convergent molecular evolution. Furthermore, a C-terminal Phe-amide motif is a common feature of many types of neuropeptides [65] and further studies are now required to investigate the relationship of the putative Spnp9-derived neuropeptide and peptides with a C-terminal Phe-amide motif that have been identified in other animals.

3.8. Spnp10

Spnp10 is a 100-residue protein comprising a predicted 24-residue N-terminal signal peptide followed by a 76-residue sequence (residues 25-100) that contains putative dibasic cleavage sites (KR) at residues 59/60 and 91/92 (Fig. 2). The N-terminal region of the protein (residues 25-58) contains ten acidic residues (D or E), indicating that this part of the protein may function as an acidic spacer peptide. On this basis we propose that it is the 30-residue

682 polypeptide formed by residues 61-90 that may be a secreted bioactive neuropeptide.
683 However, residues 61-90 also include three potential monobasic cleavage sites (R), so there
684 remains the possibility that the polypeptide formed by residues 61-90 is cleaved into two or
685 more smaller bioactive neuropeptides. The C-terminal residue (90) of the polypeptide is
686 glycine, which may be a substrate for amidation.
687 The protein sequence of Spnp10 was identified by analysis of the sequences of the
688 radial nerve cDNAs RNSP-5C3 (5': EC439329.1, GI:109403352, 3': EC438259.1,
689 GI:109402282), RNSP-1M1 (5': EC439578.1, GI:109403601, 3': EC438318.1,
690 GI:109402341) and RNSP-5A23 (5': EC439390.1, GI:109403413, 3': EC438559.1,
691 GI:109402582) but Spnp10 was not found to be represented in other cDNA libraries.
692 Spnp10 was predicted from genome sequence data by the gene prediction tool
693 Gnomon (XP_001178130.1, GI:115647054). It was not predicted, however, by the GLEAN3
694 tool that was used for genome annotation [74] and therefore it has as yet not been assigned a
695 gene ID number. The Spnp10 cDNA sequence shown in Fig. S10 is a consensus sequence
696 derived from genomic and cDNA/EST sequence data.
697 If the C-terminal glycine of the 30-residue neuropeptide derived from Spnp10 is a
698 substrate for amidation, then the mature Spnp10-derived neuropeptide would be a 29-residue
699 peptide with a C-terminal Ser-amide motif. However, this peptide does not share any apparent
700 sequence similarity with neuropeptides or peptide hormones identified in other phyla.
701
702 *3.9. Spnp11*
703 Spnp11 is a 103-residue protein comprising a predicted 21-residue N-terminal signal
704 peptide followed by a 82-residue sequence (residues 22-103) that contains a putative dibasic
705 cleavage site (KR) at residues 48 and 49 (Fig. 2). The N-terminal region of the protein
706 (residues 22-47) contains seven acidic residues (D or E), indicating that this part of the protein

may function as an acidic spacer peptide. We propose that it is the 54-residue polypeptide formed by residues 50-103 that may be a secreted bioactive neuropeptide. It is noteworthy that the 54-residue sequence includes six cysteine residues located at positions 57, 61, 64, 75, 79 and 95 because this suggests the presence of up to three potential disulphide bridges that would confer tertiary structure on the polypeptide. Alternatively a homodimeric protein could be formed by up to six intermolecular disulphide bridges. It should also be noted, however, that the 54 residue sequence also includes two potential monobasic cleavage sites (R), so there remains the possibility that the polypeptide formed by residues 50-103 is cleaved into two or more smaller bioactive neuropeptides.

The protein sequence of Spnp11 was identified by analysis of the sequences of the radial nerve cDNAs RNSP-9C11 (5': EC439009.1, GI:109403032, 3': EC437677.1, GI:109401700), RNSP-9C20 (5': EC438791.1, GI:109402814, 3': EC437882.1, GI:109401905) but Spnp11 was not found to be represented in other cDNA libraries. Spnp11 was predicted from genome sequence data by the gene prediction tool Gnomon (XP_001175484.1, GI:115666438). It was not predicted, however, by the GLEAN3 tool that was used for genome annotation [74] and therefore it has as yet not been assigned a gene ID number. The Spnp11 cDNA sequence shown in Fig. S11 is a consensus sequence derived from genomic and cDNA/EST sequence data.

The putative fifty-four residue neuropeptide derived from Spnp11 does not exhibit any apparent primary amino acid sequence similarity with neuropeptides identified in other animals. However, neuropeptides of a similar size and with six cysteine residues have been identified in other animals. For example, molt-inhibiting hormone (MIH) is a seventy-eight residue neuropeptide in the crustacean *Carcinus maenas* with six cysteine residues that form three intramolecular disulphide bonds [86].

732 *3.10. Spnp12*

733 Spnp12 is a 104-residue protein comprising a predicted 25-residue N-terminal signal
734 peptide followed by a 79-residue sequence (residues 26-104) that contains a putative dibasic
735 cleavage site (KR) at residues 41 and 42 (Fig. 2). The C-terminal region of the protein
736 (residues 43-104) contains eleven acidic residues (D or E), indicating that this part of the
737 protein may function as an acidic spacer peptide. We propose that it is the 15-residue peptide
738 (HNTFSFKGRSRYFPG) formed by residues 26-40 that may be a secreted bioactive
739 neuropeptide. The presence of a C-terminal glycine residue suggests that this peptide may be
740 amidated at the C-terminus. However, the 15-residue peptide sequence contains two potential
741 monobasic cleavage sites (R), so there remains the possibility that the peptide formed by
742 residues 26-40 is cleaved into two or more smaller bioactive neuropeptides.

743 The protein sequence of Spnp12 was identified by analysis of the sequence of the
744 radial nerve cDNA RNSP-1D20 (5': EC439240.1, GI:109403263, 3': EC438177.1,
745 GI:109402200). However, a cDNA encoding Spnp11 is also represented amongst cDNAs
746 from larvae (MPMGp691I24108, 5': CD297038.1, GI:34748115). Spnp12 was predicted
747 from genome sequence data by the gene prediction tool Gnomon (XP_001178129.1
748 GI:115620334) but Spnp12 was not, however, predicted by GLEAN3 [74] and therefore it
749 was not assigned a gene ID number during genome annotation. The Spnp12 cDNA sequence
750 shown in Fig. S12 is a consensus sequence derived from genomic and cDNA/EST sequence
751 data.

752 The putative C-terminally amidated 14-residue peptide derived from Spnp12 does not
753 share any apparent sequence similarity with neuropeptides or peptide hormones identified in
754 other phyla.

755

756 *3.11. Spnp13*

Spnp13 is a 102-residue protein comprising a predicted 18-residue N-terminal signal peptide and the putative neuropeptide LPANLARE (residues 19-26), which is bounded C-terminally by a putative dibasic cleavage site (RR) (Fig. 2). There are no other dibasic sites from residue 29 to 102, but there are potential monobasic sites (R) in this part of the protein so it is possible that other neuropeptides are derived from Spnp13. It is also noteworthy that there are nine acidic residues (D or E) in the C-terminal region of the protein from residue 64 to 102, indicating that this part of the protein may function as an acidic spacer peptide.

The protein sequence of Spnp13 was identified by analysis of the sequence of the radial nerve cDNAs RNSP-1N21 (5': EC439560.1, GI:109403583, 3': EC438178.1, GI:109402201) and RNSP-1O7 (5': EC439274.1, GI:109403297, 3': EC438582.1, GI: 109402605) but Spnp13 was not found to be represented in other cDNA libraries. Spnp13 was predicted from genome sequence data by the gene prediction tool Gnomon (XP_001176371.1, GI:115660734) but Spnp13 was not, however, predicted by GLEAN3 [74] and therefore it was not assigned a gene ID number during genome annotation. The Spnp13 cDNA sequence shown in Fig. S13 is a consensus sequence derived from genomic and cDNA/EST sequence data.

Importantly, the existence of the putative neuropeptide derived from Spnp13 (LPANLARE) has been confirmed by mass spectrometric analysis of nerve extracts from *Strongylocentrotus purpuratus* [58]. However, this peptide does not share any apparent sequence similarity with neuropeptides or peptide hormones identified in other phyla.

3.12. Spnp14

Spnp14 is a 113-residue protein comprising a predicted 26-residue N-terminal signal peptide followed by a 87-residue sequence (residues 27-113) that contains a putative dibasic cleavage site (KR) at residues 85 and 86 (Fig. 2). The N-terminal region of the protein

782 (residues 27-84) contains ten acidic residues (D or E), indicating that this part of the protein
783 may function as an acidic spacer peptide. We propose that it is the 27-residue peptide
784 (SRSGRKLRFCMDVIRNTWRLCRNTRSN) formed by residues 87-113 that may be a
785 secreted bioactive neuropeptide. The presence of two cysteine residues (underlined above)
786 suggests the presence of a disulphide bridge. Alternatively a homodimeric protein could be
787 formed by up to two intermolecular disulphide bridges.
788 The protein sequence of Spnp14 was identified by analysis of the sequence of the
789 radial nerve cDNA RNSP-9C12 (5': EC438737.1, GI:109402760, 3': EC437833.1,
790 GI:109401856) but Spnp14 was not found to be represented in other cDNA libraries. Spnp14
791 was predicted from genome sequence data by the gene prediction tool Gnomon
792 (XP_001179912.1, GI:115958765) but Spnp14 was not, however, predicted by GLEAN3
793 [74] and therefore it was not assigned a gene ID number during genome annotation. The
794 Spnp14 cDNA sequence shown in Fig. S14 is a consensus sequence derived from genomic
795 and cDNA/EST sequence data.
796 The putative twenty-seven residue neuropeptide derived from Spnp14 does not exhibit
797 any apparent primary amino acid sequence similarity with neuropeptides identified in other
798 animals. However, neuropeptides with two cysteine residues have been identified in other
799 animals. For example, the neurohypophyseal hormones vasopressin and oxytocin have two
800 cysteine residues, which form a single intramolecular disulphide bond [19].
801
802 *3.13. Spnp15*
803 Spnp15 is a 115-residue protein comprising a predicted 22-residue N-terminal signal
804 peptide followed by a 93-residue sequence (residues 23-115) that contains a putative dibasic
805 cleavage site (RR) at residues 74 and 75 (Fig. 2). The N-terminal region of the protein
806 (residues 23-73) contains fourteen acidic residues (D or E), indicating that this part of the

807 protein may function as an acidic spacer peptide. We propose that it is the 40-residue peptide
808 formed by residues 74-115 that may be a secreted bioactive neuropeptide. The presence of six
809 cysteine residues in the 40-residue polypeptide suggests that there may be up to three
810 intramolecular disulphide bridges. Alternatively a homodimeric protein could be formed by
811 up to six intermolecular disulphide bridges.
812 The protein sequence of Spnp15 was identified by analysis of the sequence of the
813 radial nerve cDNAs RNSP-9F4 (5': EC439029.1, GI:109403052, 3': EC437791.1, GI:
814 109401814), RNSP-9O10 (5': EC438734.1, GI:109402757, 3': EC437829.1, GI:109401852)
815 and RNSP-5A10 (5': EC439227.1, GI:109403250, 3': EC438016.1, GI:109402039) but
816 Spnp15 was not found to be represented in other cDNA libraries. Spnp15 was predicted from
817 genome sequence data by the gene prediction tool Gnomon (XP_001175507.1,
818 GI:115920974) but Spnp15 was not, however, predicted by GLEAN3 [74] and therefore it
819 was not assigned a gene ID number during genome annotation. The Spnp15 cDNA sequence
820 shown in Fig. S15 is a consensus sequence derived from genomic and cDNA/EST sequence
821 data.
822 The putative forty-residue neuropeptide derived from Spnp11 does not exhibit any
823 apparent primary amino acid sequence similarity with neuropeptides identified in other
824 animals. However, neuropeptides of a similar size and with six cysteine residues have been
825 identified in other animals. For example, trissin is a twenty eight-residue neuropeptide in
826 *Drosophila melanogaster* with six cysteine residues that form three intramolecular disulphide
827 bonds [38].
828
829 *3.14. Spnp16*
830 Spnp16 is a 119-residue protein comprising a predicted 20-residue N-terminal signal
831 peptide followed by a 99-residue sequence (residues 21-119) that contains a putative dibasic

832 cleavage site (KR) at residues 94 and 95 (Fig. 2). The N-terminal region of the protein
833 (residues 21-93) contains seventeen acidic residues (D or E), indicating that this part of the
834 protein may function as an acidic spacer peptide. We propose that it is the 24-residue peptide
835 (GRRPARKICINDIWKGRGGGLRCN) formed by residues 96-119 that may be a secreted
836 bioactive neuropeptide. The presence of two cysteine residues (underlined above) suggests the
837 presence of an intramolecular disulphide bridge or alternatively two intermolecular disulphide
838 bridges could form a homodimeric construct.
839 The protein sequence of Spnp16 was identified by analysis of the sequence of the
840 radial nerve cDNAs RNSP-1O8 (5': EC439410.1, GI:109403433, 3': EC438414.1,
841 GI:109402437) and RNSP-1P22 (5': EC439044.1, GI:109403067; 3': EC438269
842 GI:109402292) but Spnp16 was not found to be represented in other cDNA libraries. Spnp16
843 was predicted from genome sequence data by the gene prediction tool Gnomon
844 (XP_001176809.1, GI:115898497) but Spnp16 was not, however, predicted by GLEAN3 [74]
845 and therefore it was not assigned a gene ID number during genome annotation. The Spnp16
846 cDNA sequence shown in Fig. S16 is a consensus sequence derived from genomic and
847 cDNA/EST sequence data.
848 The putative twenty four-residue neuropeptide derived from Spnp16 does not
849 exhibit any apparent primary amino acid sequence similarity with neuropeptides identified in
850 other animals. However, as highlighted above for Spnp14, neuropeptides with two cysteine
851 residues that form a single intramolecular disulphide bond have been identified in other
852 animals (e.g. vasopressin and oxytocin [19]).
853
854 *3.16. Spnp17*
855 Spnp17 is a 120-residue protein comprising a predicted 24-residue N-terminal signal
856 peptide followed by a 96-residue sequence (residues 25-120) that contains putative dibasic

857 cleavage sites at residues 88/89 (RR) and 96/97 (RR) (Fig. 2). However, the neuropeptide
858 products of this protein are difficult to predict. If the arginine residue at position 49 is used as
859 a monobasic cleavage site, a peptide
860 (SVLKLMKYEILLKLMNDLCDELDMCPPSQVPARQAPVV) with two cysteine residues
861 (underlined) would be liberated, with the potential for an intramolecular disulphide bridge or
862 alternatively two intermolecular disulphide bridges giving rise to a homodimeric construct. A
863 potential second neuropeptide (RGGAHLFWRTGVLNKSPIMKAAN) could be liberated
864 from the protein if the dibasic cleavage site at residues 96/97 is used. It is noteworthy that in
865 the N-terminal part of the protein following the signal peptide (residues 25-61) there are seven
866 acidic residues (D or E), indicating that this part of the protein may function as an acidic
867 spacer peptide.
868 The protein sequence of Spnp17 was identified by analysis of the sequence of the
869 radial nerve cDNA RNSP-5E13 (5': EC439292.1, GI:109403315, 3': EC438502.1
870 GI:109402525) but Spnp17 was also found to be represented in a lantern cDNA library (LSP-
871 2M15, 5': EC435368.1, GI:109399391). Spnp17 was not predicted from genome sequence
872 data by the gene prediction tools Gnomon or GLEAN3 and therefore it was not assigned a
873 gene ID number during genome annotation [74]. The Spnp17 cDNA sequence shown in Fig.
874 S17 is a consensus sequence derived from genomic and cDNA/EST sequence data.
875 The putative neuropeptides derived from Spnp17 do not exhibit any apparent
876 primary amino acid sequence similarity with neuropeptides identified in other animals.
877
878 *3.17. Spnp18*
879 Spnp18 is a 121-residue protein comprising a predicted 24-residue N-terminal signal
880 peptide followed by a 97-residue sequence (residues 25-121) that contains putative dibasic
881 cleavage sites (KR) at residues 38/39 and 110/111 (Fig. 2). We propose that it is the 70-

882 residue polypeptide formed by residues 40-109 that may be a secreted bioactive neuropeptide.
883 It is noteworthy that this putative 70-residue neuropeptide contains eight cysteine residues,
884 which may form up to four intramolecular disulphide bridges. Alternatively, up to eight
885 intermolecular disulphide bridges may give rise to dimeric constructs of the polypeptide.
886 The protein sequence of Spnp18 was identified by analysis of the sequence of the
887 radial nerve cDNAs RNSP-1M16 (5': EC439524.1, GI:109403547, 3': EC438421.1,
888 GI:109402444), RNSP-1O1 (5': EC439579.1, GI:109403602, 3': EC438322.1,
889 GI:109402345), RNSP-9M3 (5': EC438932.1, GI:109402955, 3': EC437708.1,
890 GI:109401731), RNSP-1A12 (5': EC439372.1, GI:109403395; 3': EC438395.1,
891 GI:109402418), RNSP-1E15 (5': EC439205.1, GI:109403228, 3': EC438375.1,
892 GI:109402398), RNSP-1I12 (5': EC439376.1, GI:109403399, 3': EC438367.1,
893 GI:109402390) and RNSP-9L6 (5': EC438939.1, GI:109402962, 3': EC437751.1,
894 GI:109401774). Spnp18 was not represented in other cDNA libraries. Spnp18 was predicted
895 from genome sequence data by the gene prediction tool Gnomon (XP_001175944.1,
896 GI:115839524) but Spnp18 was not, however, predicted by GLEAN3 [74] and therefore it
897 was not assigned a gene ID number during genome annotation. The Spnp18 cDNA sequence
898 shown in Fig. S18 is a consensus sequence derived from genomic and cDNA/EST sequence
899 data.
900 The putative seventy-residue neuropeptide derived from Spnp18 does not exhibit any
901 apparent primary amino acid sequence similarity with neuropeptides identified in other
902 animals, which would be indicative of a common evolutionary relationship. However,
903 neuropeptides of a similar size and with eight cysteine residues have been identified in other
904 animals. For example, schistosomin is a seventy-nine residue anti-gonadotropic peptide in the
905 pond snail *Lymnaea stagnalis* and it has eight cysteine residue that are thought to form four
906 intramolecular disulphide bonds [36].

907

908 *3.18. Spnp19*

909 Spnp19 is a 129-residue protein comprising a predicted 22-residue N-terminal signal
910 peptide followed by a 107-residue sequence (residues 23-129) that contains putative dibasic
911 cleavage sites (KR) at residues 55/56 and 122/123 (Fig. 2). We propose that it is the 65-
912 residue polypeptide formed by residues 57-121 that may be a secreted bioactive neuropeptide.
913 It is noteworthy that this putative 65-residue neuropeptide contains two cysteine residues
914 (positions 54 and 65 in the putative peptide), which may form an intramolecular disulphide
915 bridge. Alternatively, two intermolecular disulphide bridges may give rise to dimeric
916 constructs of the polypeptide.

917 The protein sequence of Spnp19 was identified by analysis of the sequence of the
918 radial nerve cDNA RNSP-9F9 (5': EC438819.1, GI:109402842, 3': EC437908.1,
919 GI:109401931) and was not represented in other cDNA libraries. Spnp19 was predicted from
920 genome sequence data by the gene prediction tool Gnomon (XP_001176669.1,
921 GI:115722995) but Spnp19 was not, however, predicted by GLEAN3 [74] and therefore it
922 was not assigned a gene ID number during genome annotation. The Spnp19 cDNA sequence
923 shown in Fig. S19 is a consensus sequence derived from genomic and cDNA/EST sequence
924 data.

925 The putative sixty five-residue neuropeptide derived from Spnp16 does not exhibit
926 any apparent primary amino acid sequence similarity with neuropeptides identified in other
927 animals.

928

929 *3.19. Spnp20*

930 Spnp20 is a 157-residue protein comprising a predicted 22-residue N-terminal signal
931 peptide followed by a 135-residue sequence (residues 23-157) that contains a putative dibasic

cleavage site (RR) at residues 112/113 (Fig. 2). We propose that it is the 44-residue polypeptide formed by residues 114-157 that may be a secreted bioactive neuropeptide. It is noteworthy that this putative 44-residue neuropeptide contains two cysteine residues (positions 18 and 24 in the putative peptide), which may form an intramolecular disulphide bridge. Alternatively, two intermolecular disulphide bridges may give rise to dimeric constructs of the polypeptide.

The protein sequence of Spnp20 was identified by analysis of the sequence of the radial nerve cDNA RNSP-1I6 (5': EC439447.1, GI:109403470, 3': EC438404.1, GI:109402427) but it is also represented in many other cDNA libraries, including unfertilised egg (e.g. MPMGp621P0242, 5': CD316932.1, GI:34788993), 7 hour cleavage stage (e.g. CALTp538D011, 5': CD319009.1, GI:34791070, 3': CD290037.1, GI:34741114), 20 hour blastula stage (e.g. CALTp537G0419, 5': CD336553.1, GI:34803079, 3': CD324544.1, GI:34796605), primary mesenchyme cells (e.g. PMCSPR2-160F9, 5': DN585364.1, GI:61235578, 3': DN568702.1, GI:61127741) and larvae (e.g. MPMGp691B14106, 5': CD307438.1, GI:34752487). Spnp20 was predicted from automated analysis of genomic sequence data by gene prediction tools (Gnomon - XP_799788.2, GI:390348447; GLEAN3_14142) and assigned the gene ID number SPU_014142 [74]. The Spnp20 cDNA sequence shown in Fig. S20 is a consensus sequence derived from genomic, cDNA/EST and also RNAseq (WHL22.545917.1) sequence data.

The putative forty four-residue neuropeptide derived from Spnp20 does not exhibit any apparent primary amino acid sequence similarity with neuropeptides identified in other animals.

4.8. Conclusions

The identification of precursor proteins for putative neuropeptides in the sea urchin *Strongylocentrotus purpuratus*, as reported here, is of interest from two perspectives.

Firstly, it contributes to a growing body of comparative data on neuropeptides, providing new insights on the phylogenetic distribution and evolutionary origins of neuropeptide families in the animal kingdom. For example, an important finding from this study is the discovery that calcitonin-like peptides with two N-terminally located cysteine residues are found not only in chordates but also in a non-chordate deuterostome and therefore the origin of this type of peptide can be traced back to the common ancestor of extant deuterostomes. Additionally, the discovery of pedal peptide-like neuropeptides in *Strongylocentrotus purpuratus* has revealed a bilaterian family of pedal peptide/orcokinin-type neuropeptides.

Secondly, discovery of twenty putative neuropeptide precursors provides a solid foundation for a comprehensive investigation of neuropeptide function in a model echinoderm. There are many fascinating aspects of echinoderm biology, including remarkable powers of regeneration [78] following autotomy of body parts [88] and the "mutability" of echinoderm connective tissue [89]. There is evidence that neuropeptides are important regulators of these and many other aspects of echinoderm biology [7, 22] and the putative neuropeptides identified here in *Strongylocentrotus purpuratus* provide material for experimental studies on sea urchins. Moreover, *Strongylocentrotus purpuratus* is first and foremost a model system for development biology [55] and the neuropeptide precursors identified here provide material for developmental analysis of neuropeptide expression and function and analysis of the organisation of neuropeptide systems in the simple nervous system of the free-swimming larval stage of this species.

Acknowledgements

We are grateful to Eric Davidson, Andy Cameron and colleagues at Caltech (USA) and to Erica Sodergren and colleagues at Baylor College of Medicine Human Genome Sequencing Centre (USA) for generating and sequencing the *Strongylocentrotus purpuratus* radial nerve cDNAs (GI:109401590 – GI:109403615) that were analysed in this study. The work reported here was supported by a grant from the University of London Central Research Fund. The support of Chris and Lyndon Rowe is also gratefully acknowledged. We thank Paola Oliveri (University College London) for her helpful feedback during the preparation of this paper.

Figure legends

Figure 1.

Amino acid sequences of the *Strongylocentrotus purpuratus* putative neuropeptide precursors Spnp1 – Spnp9. The predicted N-terminal signal peptide for each precursor is shown in bold lettering. Putative neuropeptides are shown in white with black highlighting, with the exception of cysteine residues, which are shown in white with light grey highlighting. C-terminal glycine residues that are putative substrates for amidation are shown in white with dark grey highlighting. Putative cleavage sites are shown with black letters and light grey highlighting.

Figure 2.

Amino acid sequences of the *Strongylocentrotus purpuratus* putative neuropeptide precursors Spnp10 – Spnp20. The predicted N-terminal signal peptide for each precursor is shown in bold lettering. Putative neuropeptides are shown in white with black highlighting, with the exception of cysteine residues, which are shown in white with light grey highlighting. C-terminal glycine residues that are putative substrates for amidation are shown in white with dark grey highlighting. Putative cleavage sites are shown with black letters and light grey highlighting.

Figure 3.

Sequence alignment of the *Strongylocentrotus purpuratus* calcitonin-like peptide (SpCTLP) with human calcitonin, human calcitonin gene-related peptide (CGRP), *Ciona* calcitonin (Ci CT), *Drosophila* calcitonin-like diuretic peptide (DH 31), and *Homarus* calcitonin-like diuretic peptide (DH 31). Cysteine residues are shown in white with black highlighting, the

basic amino acids Lys and Arg are shown in black with light grey highlighting and the acidic residues Glu and Asp are shown in black with dark grey highlighting. All other amino acids are classified as hydrophobic (white with light grey highlighting) or hydrophilic (white with dark grey highlighting). C-terminal amide groups are shown as a lowercase "a". Note that SpCTLP has two cysteine residues in its N-terminal region, a character that it shares with human calcitonin, human CGRP and *Ciona* CT but not with calcitonin-like diuretic peptides in arthropods (*Drosophila* DH 31 and *Homarus* DH31). Therefore, this may be a conserved and characteristic feature of calcitonin-type peptides in deuterostomes. References: 1. GI:179820, 2. GI:76880478, 3. GI:283046319, 4. This paper, 5. GI:17647327, 6. GI: 260594183 [13]

Figure 4.

Sequence alignment of *Strongylocentrotus purpuratus* pedal peptide-like neuropeptides SpPPLN1d and SpPPLN2h with pedal peptides from the mollusc *Aplysia californica*, a pedal peptide-like neuropeptide derived from the "FDSIG" precursor in the annelid *Platynereis dumerilii*, orcokinin-type neuropeptides in the crustacean *Procambrus clarkii* and the insect *Nasonia vitripennis* and pedal peptide/orcokinin-like peptides derived from the NPL14 and NPL15 precursor proteins in *Caenorhabditis elegans*. The basic amino acids Lys and Arg are shown in black with light grey highlighting and the acidic residues Glu and Asp are shown in black with dark grey highlighting. All other amino acids are classified as hydrophobic (white with light grey highlighting) or hydrophilic (white with dark grey highlighting). References: 1. This paper, 2. GI:325297152, GI: 325296771, GI: 325296775 [60]. 3. GI: 332167919 [14], 4. GI:392926792, GI:7498042, 5. GI:38258254 [90], 6. GI: 345489156.

References

[1] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, et al., The genome sequence of *Drosophila melanogaster*. Science. 287 (2000) 2185-2195.

[2] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. J Mol Biol. 215 (1990) 403-410.

[3] S.G. Amara, V. Jonas, M.G. Rosenfeld, E.S. Ong, R.M. Evans, Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature. 298 (1982) 240-244.

[4] M. Anctil, Chemical transmission in the sea anemone *Nematostella vectensis*: A genomic perspective. Comp Biochem Physiol Part D Genomics Proteomics. 4 (2009) 268-289.

[5] L. Avery, *Caenorhabditis elegans* behavioral genetics: where are the knobs? BMC Biol. 8 (2010) 69.

[6] J.D. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340 (2004) 783-795.

[7] R. Birenheide, M. Tamori, T. Motokawa, M. Ohtani, E. Iwakoshi, Y. Muneoka, et al., Peptides controlling stiffness of connective tissue in sea cucumbers. Biol Bull. 194 (1998) 253-259.

[8] J.E. Blair, S.B. Hedges, Molecular phylogeny and divergence times of deuterostome animals. Mol Biol Evol. 22 (2005) 2275-2284.

[9] S.J. Bourlat, T. Juliusdottir, C.J. Lowe, R. Freeman, J. Aronowicz, M. Kirschner, et al., Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature. 444 (2006) 85-88.

[10] R.D. Burke, L.M. Angerer, M.R. Elphick, G.W. Humphrey, S. Yaguchi, T. Kiyama, et al., A genomic view of the sea urchin nervous system. Dev Biol. 300 (2006) 434-460.

[11] R.A. Cameron, M. Samanta, A. Yuan, D. He, E. Davidson, SpBase: the sea urchin genome database and web site. Nucleic Acids Res. 37 (2009) D750-754.

[12] A.B. Chaet, R.A. McConnaughy, Physiologic activity of nerve extracts. Biol Bull. 117 (1959) 407.

[13] A.E. Christie, J.S. Stevens, M.R. Bowers, M.C. Chapline, D.A. Jensen, K.M. Schegg, et al., Identification of a calcitonin-like diuretic hormone that functions as an intrinsic modulator of the American lobster, *Homarus americanus*, cardiac neuromuscular system. The Journal of experimental biology. 213 (2010) 118-127.

[14] M. Conzelmann, S.L. Offenburger, A. Asadulina, T. Keller, T.A. Munch, G. Jekely, Neuropeptides regulate swimming depth of *Platynereis* larvae. P Natl Acad Sci USA. 108 (2011) E1174-1183.

[15] D. de Wied, M. Diamant, M. Fodor, Central nervous system effects of the neurohypophyseal hormones and related peptides. Front Neuroendocrinol. 14 (1993) 251-302.

[16] P. Dehal, Y. Satou, R.K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, et al., The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science. 298 (2002) 2157-2167.

[17] G.J. Dockray, J.R. Reeve, Jr., J. Shively, R.J. Gayton, C.S. Barnard, A novel active pentapeptide from chicken brain identified by antibodies to FMRFamide. Nature. 305 (1983) 328-330.

[18] G.J. Dockray, C. Vaillant, R.G. Williams, R.J. Gayton, N.N. Osborne, Vertebrate brain-gut peptides related to FMRFamide and Met-enkephalin Arg6Phe7. Peptides. 2 Suppl 2 (1981) 25-30.

[19] V. Du Vigneaud, Hormones of the posterior pituitary gland: oxytocin and vasopressin. Harvey Lect. 50 (1954) 1-26.
[20] B.A. Eipper, D.A. Stoffers, R.E. Mains, The biosynthesis of neuropeptides: peptide alpha-amidation. Annu Rev Neurosci. 15 (1992) 57-85.
[21] M.R. Elphick, The Protein Precursors of Peptides That Affect the Mechanics of Connective Tissue and/or Muscle in the Echinoderm *Apostichopus japonicus*. PLoS One. 7 (2012) e44492.
[22] M.R. Elphick, R. Melarange, Neural control of muscle relaxation in echinoderms. J Exp Biol. 204 (2001) 875-885.
[23] M.R. Elphick, D.A. Price, T.D. Lee, M.C. Thorndyke, The SALMFamides: a new family of neuropeptides isolated from an echinoderm. Proc Biol Sci. 243 (1991) 121-127.
[24] M.R. Elphick, J.R. Reeve, Jr., R.D. Burke, M.C. Thorndyke, Isolation of the neuropeptide SALMFamide-1 from starfish using a new antiserum. Peptides. 12 (1991) 455-459.
[25] M.R. Elphick, M.L. Rowe, NGFFFamide and echinotocin: structurally unrelated myoactive neuropeptides derived from neurophysin-containing precursors in sea urchins. The Journal of experimental biology. 212 (2009) 1067-1077.
[26] M.R. Elphick, M.C. Thorndyke, Molecular characterisation of SALMFamide neuropeptides in sea urchins. The Journal of experimental biology. 208 (2005) 4273-4282.
[27] W.H. Fischer, J. Spiess, Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. P Natl Acad Sci USA. 84 (1987) 3628-3632.
[28] R.M. Freeman, Jr., M. Wu, M.M. Cordonnier-Pratt, L.H. Pratt, C.E. Gruber, M. Smith, et al., cDNA sequences for transcription factors and signaling proteins of the hemichordate *Saccoglossus kowalevskii*: efficacy of the expressed sequence tag (EST) approach for evolutionary and developmental studies of a new organism. The Biological bulletin. 214 (2008) 284-302.
[29] L.D. Fricker, J. Lim, H. Pan, F.Y. Che, Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev. 25 (2006) 327-344.
[30] K. Furuya, R.J. Milchak, K.M. Schegg, J. Zhang, S.S. Tobe, G.M. Coast, et al., Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. P Natl Acad Sci USA. 97 (2000) 6469-6474.
[31] A. Gorbman, S.A. Sower, Evolution of the role of GnRH in animal (Metazoan) biology. Gen Comp Endocr. 134 (2003) 207-213.
[32] M. Hamada, N. Shimozono, N. Ohta, Y. Satou, T. Horie, T. Kawada, et al., Expression of neuropeptide- and hormone-encoding genes in the *Ciona intestinalis* larval brain. Developmental biology. 352 (2011) 202-214.
[33] F. Hauser, M. Williamson, G. Cazzamali, C.J. Grimmelikhuijzen, Identifying neuropeptide and protein hormone receptors in *Drosophila melanogaster* by exploiting genomic data. Brief Funct Genomic Proteomic. 4 (2006) 321-330.
[34] R.S. Hewes, P.H. Taghert, Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. Genome Res. 11 (2001) 1126-1142.
[35] L. Holden-Dye, R.J. Walker, The roles of neuropeptides in *Caenorhabditis elegans* including their importance in the regulation of feeding and metabolism. Protein Pept Lett (2012).
[36] P.L. Hordijk, H.D. Schallig, R.H. Ebberink, M. de Jong-Brink, J. Joosse, Primary structure and origin of schistosomin, an anti-gonadotropic neuropeptide of the pond snail *Lymnaea stagnalis*. Biochem J. 279 (Pt 3) (1991) 837-842.

[37] Y.S. Hsieh, M.C. Fann, W.C. Wan, P.S. Wang, The presence of a thyrotropin-releasing hormone-like factor in the grass shrimp (black tiger prawn, *Penaeus monodon*). Chin J Physiol. 33 (1990) 179-190.
[38] T. Ida, T. Takahashi, H. Tominaga, T. Sato, K. Kume, K. Yoshizawa-Kumagaye, et al., Identification of the endogenous cysteine-rich peptide trissin, a ligand for an orphan G protein-coupled receptor in *Drosophila*. Biochem Biophys Res Commun. 414 (2011) 44-48.
[39] M. Inoue, R. Birenheide, O. Koizumi, Y. Kobayakawa, Y. Muneoka, T. Motokawa, Localization of the neuropeptide NGIWYamide in the holothurian nervous system and its effects on muscular contraction. Proc Biol Sci. 266 (1999).
[40] E. Iwakoshi, M. Ohtani, T. Takahashi, Y. Muneoka, T. Ikeda, T. Fujita, et al., Comparative aspects of structure and action of bioactive peptides isolated from the sea cucumber *Stichopus japonicus*. In: M. Ohno, (Ed.), Peptide Chemistry 1994, Protein Research Foundation, Osaka, 1995, pp. 261-264.
[41] E. Iwakoshi, K. Takuwa-Kuroda, Y. Fujisawa, M. Hisada, K. Ukena, K. Tsutsui, et al., Isolation and characterization of a GnRH-like peptide from *Octopus vulgaris*. Biochem Biophys Res Commun. 291 (2002) 1187-1193.
[42] J. Joosse, What is special about peptides as neuronal messengers? In: M.C. Thorndyke, G.J. Goldsworthy, (Eds.), Neurohormones in invertebrates, Cambridge University Press, Cambridge, 1988, pp. 1-3.
[43] H. Kanatani, Hormones in echinoderms. In: E.J.W. Barrington, (Ed.), Hormones and Evolution, Academic Press, London, 1979, pp. 273-307.
[44] S. Kato, S. Tsurumaru, M. Taga, T. Yamane, Y. Shibata, K. Ohno, et al., Neuronal peptides induce oocyte maturation and gamete spawning of sea cucumber, *Apostichopus japonicus*. Developmental biology. 326 (2009) 169-176.
[45] T. Kawada, T. Sekiguchi, T. Sakai, M. Aoyama, H. Satake, Neuropeptides, hormone peptides, and their receptors in *Ciona intestinalis*: an update. Zool Sci. 27 (2010) 134-153.
[46] D.K. Kim, E.B. Cho, M.J. Moon, S. Park, J.I. Hwang, O. Kah, et al., Revisiting the evolution of gonadotropin-releasing hormones and their receptors in vertebrates: secrets hidden in genomes. Gen Comp Endocr. 170 (2011) 68-78.
[47] C. Li, K. Kim, Neuropeptide gene families in *Caenorhabditis elegans*. Adv Exp Med Biol. 692 (2010) 98-137.
[48] C. Li, L.S. Nelson, K. Kim, A. Nathoo, A.C. Hart, Neuropeptide gene families in the nematode *Caenorhabditis elegans*. Ann Ny Acad Sci. 897 (1999) 239-252.
[49] M. Lindemans, T. Janssen, I. Beets, L. Temmerman, E. Meelkop, L. Schoofs, Gonadotropin-releasing hormone and adipokinetic hormone signaling systems share a common evolutionary origin. Front Endocrinol (Lausanne). 2 (2011) 16.
[50] M. Lindemans, F. Liu, T. Janssen, S.J. Husson, I. Mertens, G. Gade, et al., Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*. P Natl Acad Sci USA. 106 (2009) 1642-1647.
[51] P.E. Lloyd, C.M. Connolly, Sequence of pedal peptide: a novel neuropeptide from the central nervous system of *Aplysia*. Journal of neuroscience. 9 (1989) 312-317.
[52] G.O. Mackie, On the Visceral Nervous-System of *Ciona*. JOURNAL OF THE MARINE BIOLOGICAL ASSOCIATION OF THE UNITED KINGDOM. 75 (1995) 141-151.
[53] N.J. Marks, A.G. Maule, Neuropeptides in helminths: occurrence and distribution. Adv Exp Med Biol. 692 (2010) 49-77.

[54] C. McCall, T. Singer, The animal and human neuroendocrinology of social cognition, motivation and behavior. Nat Neurosci. 15 (2012) 681-688.
[55] D.R. McClay, Evolutionary crossroads in developmental biology: sea urchins. Development. 138 (2011) 2639-2648.
[56] T. Meeusen, I. Mertens, E. Clynen, G. Baggerman, R. Nichols, R.J. Nachman, et al., Identification in *Drosophila melanogaster* of the invertebrate G protein-coupled FMRFamide receptor. P Natl Acad Sci USA. 99 (2002) 15363-15368.
[57] R. Melarange, M.R. Elphick, Comparative analysis of nitric oxide and SALMFamide neuropeptides as general muscle relaxants in starfish. J Exp Biol. 206 (2003) 893-899.
[58] G. Menschaert, T.T. Vandekerckhove, G. Baggerman, B. Landuyt, J.V. Sweedler, L. Schoofs, et al., A hybrid, de novo based, genome-wide database search approach applied to the sea urchin neuropeptidome. J Proteome Res. 9 (2010) 990-996.
[59] M. Mita, M. Yoshikuni, K. Ohno, Y. Shibata, B. Paul-Prasanth, S. Pitchayawasin, et al., A relaxin-like peptide purified from radial nerves induces oocyte maturation and ovulation in the starfish, *Asterina pectinifera*. P Natl Acad Sci USA. 106 (2009) 9507-9512.
[60] L.L. Moroz, J.R. Edwards, S.V. Puthanveettil, A.B. Kohn, T. Ha, A. Heyland, et al., Neuronal transcriptome of *Aplysia*: neuronal compartments and circuitry. Cell. 127 (2006) 1453-1467.
[61] D.R. Nassel, A.M. Winther, Drosophila neuropeptides in regulation of physiology and behavior. Prog Neurobiol. 92 (2010) 42-104.
[62] M. Ogoshi, K. Inoue, K. Naruse, Y. Takei, Evolutionary history of the calcitonin gene-related peptide family in vertebrates revealed by comparative genomic analyses. Peptides. 27 (2006) 3154-3164.
[63] G. Pesole, S. Liuni, G. Grillo, C. Saccone, Structural and compositional features of untranslated regions of eukaryotic mRNAs. Gene. 205 (1997) 95-102.
[64] H. Philippe, H. Brinkmann, R.R. Copley, L.L. Moroz, H. Nakano, A.J. Poustka, et al., Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature. 470 (2011) 255-258.
[65] D.A. Price, M.J. Greenberg, The Hunting of the FaRPs: The Distribution of FMRFamide-Related Peptides. The Biological Bulletin. 177 (1989) 198-205.
[66] N.H. Putnam, T. Butts, D.E. Ferrier, R.F. Furlong, U. Hellsten, T. Kawashima, et al., The amphioxus genome and the evolution of the chordate karyotype. Nature. 453 (2008) 1064-1071.
[67] G.J. Roch, E.R. Busby, N.M. Sherwood, Evolution of GnRH: diving deeper. Gen Comp Endocr. 171 (2011) 1-16.
[68] M.L. Rowe, M.R. Elphick, Discovery of a second SALMFamide gene in the sea urchin *Strongylocentrotus purpuratus* reveals that L-type and F-type SALMFamide neuropeptides coexist in an echinoderm species. Mar Genomics. 3 (2010) 91-97.
[69] A.V. Schally, A. Arimura, A.J. Kastin, H. Matsuo, Y. Baba, T.W. Redding, et al., Gonadotropin-releasing hormone: one polypeptide regulates secretion of luteinizing and follicle-stimulating hormones. Science. 173 (1971) 1036-1038.
[70] N.G. Seidah, M. Chretien, Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. Brain Res. 848 (1999) 45-62.
[71] T. Sekiguchi, T. Kawashima, Y. Satou, N. Satoh, Further EST analysis of endocrine genes that are preferentially expressed in the neural complex of *Ciona intestinalis*: receptor and enzyme genes associated with endocrine system in the neural complex. Gen Comp Endocr. 150 (2007) 233-245.

[72] T. Sekiguchi, N. Suzuki, N. Fujiwara, M. Aoyama, T. Kawada, K. Sugase, et al., Calcitonin in a protochordate, *Ciona intestinalis*--the prototype of the vertebrate calcitonin/calcitonin gene-related peptide superfamily. Febs J. 276 (2009) 4437-4447.
[73] N.M. Sherwood, J.A. Tello, G.J. Roch, Neuroendocrinology of protochordates: insights from *Ciona* genomics. Comparative biochemistry and physiology. Part A, Molecular & integrative physiology. 144 (2006) 254-271.
[74] E. Sodergren, G.M. Weinstock, E.H. Davidson, R.A. Cameron, R.A. Gibbs, R.C. Angerer, et al., The genome of the sea urchin *Strongylocentrotus purpuratus*. Science. 314 (2006) 941-952.
[75] F.L. Strand, Neuropeptides: regulators of physiological processes, The MIT Press: Cambridge, Massachusetts, 1999.
[76] K. Terakado, Induction of gamete release by gonadotropin-releasing hormone in a protochordate, *Ciona intestinalis*. Gen Comp Endocr. 124 (2001) 277-284.
[77] The *C. elegans* Sequencing Consortium, Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 282 (1998) 2012-2018.
[78] M.C. Thorndyke, W.C. Chen, P.W. Beesley, M. Patruno, Molecular approach to echinoderm regeneration. Microsc Res Tech. 55 (2001) 474-485.
[79] P.S. Tsai, Gonadotropin-releasing hormone in invertebrates: structure, function, and evolution. Gen Comp Endocr. 148 (2006) 48-53.
[80] Q. Tu, R.A. Cameron, K.C. Worley, R.A. Gibbs, E.H. Davidson, Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. Genome Res (2012).
[81] J. Vanden Broeck, Neuropeptides and their precursors in the fruitfly, *Drosophila melanogaster*. Peptides. 22 (2001) 241-254.
[82] J.A. Veenstra, Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Arch Insect Biochem Physiol. 43 (2000) 49-63.
[83] J.A. Veenstra, Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. Gen Comp Endocr. 167 (2010) 86-103.
[84] J.A. Veenstra, Neuropeptide evolution: neurohormones and neuropeptides predicted from the genomes of *Capitella teleta* and *Helobdella robusta*. Gen Comp Endocr. 171 (2011) 160-175.
[85] M. Wallis, Molecular evolution of the thyrotrophin-releasing hormone precursor in vertebrates: insights from comparative genomics. J Neuroendocrinol. 22 (2010) 608-619.
[86] S.G. Webster, Amino acid sequence of putative moult-inhibiting hormone from the crab *Carcinus maenas*. Proceedings. Biological sciences / The Royal Society. 244 (1991) 247-252.
[87] S.E. Wendelaar Bonga, P.K. Pang, Control of calcium regulating hormones in the vertebrates: parathyroid hormone, calcitonin, prolactin, and stanniocalcin. Int Rev Cytol. 128 (1991) 139-213.
[88] I.C. Wilkie, Autotomy as a prelude to regeneration in echinoderms. Microsc Res Tech. 55 (2001) 369-396.
[89] I.C. Wilkie, Mutable collagenous tissue: overview and biotechnological perspective. Prog Mol Subcell Biol. 39 (2005) 221-250.
[90] Y. Yasuda-Kamatani, A. Yasuda, Identification of orcokinin gene-related peptides in the brain of the crayfish *Procambarus clarkii* by the combination of MALDI-TOF and on-line capillary HPLC/Q-Tof mass spectrometries and molecular cloning. Gen Comp Endocr. 118 (2000) 161-172.

[91] K.A. Young, Y. Liu, Z. Wang, The neurobiology of social attachment: A comparative approach to behavioral, neuroanatomical, and neurochemical studies. Comp Biochem Physiol C Toxicol Pharmacol. 148 (2008) 401-410.
[92] S.S. Yun, M.C. Thorndyke, M.R. Elphick, Identification of novel SALMFamide neuropeptides in the starfish *Marthasterias glacialis*. Comparative biochemistry and physiology. Part A, Molecular & integrative physiology. 147 (2007) 536-542.

Figure 1

Spnp1 (L-type SALMFamide)
MQVQQITVFLVACTLSVLVVAYAQEDAETVLLNRLRDIAARAAAGELPDFFADVDDYKRGGKKNMG
SIHSHSGIHFGKRRDSESSERARNTKRMRLHPGLLFGKRAPVQKWDQWQAQDTYNPDWELGQFN

Spnp2 (SpGnRHP)
MKQIITSLVSISAALLLFVLISEYTPRCNGQVHHRFSGWRPGGKKRSDAAEVNSNKITIERPQLPI
CQTTEERQLLEGDSDILGDLRRAANRMRLLQLFNLSKTRLNDLNDATSNEVDERPVYGDYLGTGL

Spnp3 (SpTRHLP)
MWACILGYVTWGGAALPTILGKELVLSENDGPEIADWVQGKEIPLRNQYWGDVAEEEEEEELGMLS
PDSEKRQYPGGKRQYPGGKRQYPGGKRQYPGGKRQYPGGKRQFPAGKRQFVGGELIPSPELRQWPGGKRQWPG
GKRQWPGGKRQYPGGKRQYPGGKRQWPEVKRQYPGGKRSEDDQDLLPMEIRQYPGGKRQWPGGKRQ
YPGGKRQYPGGKRQFPGGKRQFVGGEALEQESNINKRFAPEDDTMDFFRLSQLYDTNDNIVADEGE
LALEDLLDDIMVDTRPEFEDPRDLLLGNVDQEDVLALDLSALLGDRNPNNGW

Spnp4 (SpCTLPP)
MKSTVIVTLTICCLLYQTTRAASLTNRDGLSRQDILDLLQLYEEPIRQEGGDKRSKGCGSFSGCMQ
MEVAKNRVAALLRNSNAHLFGLNGPGKRRRSVDDLPQVNDAETE

Spnp5 (SpANPP)
MSRNAYLWAGLLLGALCLLITTTSIKADGEVTEDVDKRANYFRGRGRKPGKRDEPDAALVPDDDLS
EDKRANMFRSRLRGKGKRDDPDAAMLPGDWDEEKRANMFRSRLRGNGKRDDPDAAMLPGDWDEEKR
ANMFRSRLRGKCKRDEPDAAEALVPGDWEEEKRANMFRSRLRGKGKRDDPDAAEALVPGDDLSEEK
RANMFRSRLRGKGKRDDPDAAEALVPGDDLSEEKRANMFRSRLRGKCKRDDPDAAEALVPGGDLSE
EKRANMFRSRLRGKCKRDDPDAAEALVPGGDLSEEKRANMFRSRLRGKGKRDDPDAAEALVPGDWD
EEKRANMFRSRLRGKGKRDDPDAALVGDDFGDEFVDEEKRANMFRSRLRGNGKRDDPDAALVDEFM
DEEKRANYFRGRGRRPGKRDEPDAALVEDEKRANFRARQRPKLGK

Spnp6 (SpPPLNP1)
MKFSGNGRGAFLVVNLIFVLCLVDHMAECRPARKTRDVDEDLEKEEDSLINALEKVLADEEVIDNA
ENDSDDETGITDRELSLMLSMLRDDVSPSRLRGYFGGKWRPAYYPSESLHVGALEPLATGFLPSRY
SGQKKRFLTGALEPLSSGFIKKGFNTGAMEPLGSGFIKKGFNSGAMEPLGAGFFKKGFNSGAMEPL
GAGFFKKGFNSGAMEPLGAGFFKKGFNSGAMEPLGAGFFKKGFNSGAMEPLGAGFFKKGFNSGAME
PLGAGFFKKGFNSGAMEPLGAGFFKKGFNSGAMEPLGSGFIKKGFNSGAMEPLGSGFIKKGFNNGA
MEPLGSGFIKKGFNSGAMEPLGSGFIKKGFNSGAMEPLGSGFIKKGFNTGAMEPLGSGFIKKGFNS
GAMEPLGSGFIKKDFNTGAMEPLGSGFIKKGFNSGAMEPLGAGFFKKGFHAGAMEPLSSGFIDGKR
GFYNGAMEPLSAGFHQGKRGFHEGEMDKDKKGFHNGAMEPLKSGFLKD

Spnp7 (SpPPLNP2)
MNNYAFLFCLACAIGQVWTLPIEDKDGLDIEDQEEAEKRFGSMNMEPLVSGFYKRFGSGLDSMQSG
FYKKNFGSGLNMEPMQSGFYKKNFGGSMEPMQSGFYKRFGGAMEPMSSGFYKRFGSGSLEPMSSGF
YKKNFGGSLEPMQSGFYKRFGGANEPMRSGFFKRFGSGSLEPMSSGFYKKNFGGSLDAMQSGFYKR
SQEETD

Spnp8
MANRQLLALAFIVSLALAVVEARNFHAAMGGPRPWQAGMKQQSALPDKGTNPFLKRLKQIVFQPDG
FYDPGMDHFAFGEAFNADE

Spnp9
MRSSLAVLLLACLAAIISRESPVQAVPRIRPAILQHGMPFGKRGYSGNNARDCFHRALNDDKNSEE
LVNLIEAWYRMKVEDGLSCMNGLSAFDEAAA

Figure 2

Spnp10
MKSVYQVVLAFLAVLVCVAWTCQAYGLDQDEYRRGAAENALDEQEIYEIIESLEHAMSKRGSVKHL
GLANVDNWRMMKNVNRLRNLLNSGKRSDQQLDSQ

Spnp11
MNSLILVVMGLLLLTAELIPAAPAPYFDEDAMDLMDPVFNFKDDSAVKRSPMLQKSCIYTCLACSK
NTQMTMPECIYGCQSAGRDPSQARAYNACHKYLHSGR

Spnp12
MDSNMTVRSLVILSVLLLAVVSCHAHNTFSFKGRSRYFPGKRAITDGSAVDTASQRFESINLDDFQ
KPESQLTLREMLTELRGYCDFLLKLLDGVRPDLPQQRK

Spnp13
MELRLFLLVVLFCALATSLPANLARERRTTNPVLRDKGRESMKTKQFRIGYRYGRAWQPPTTLDDN
VYGADNYDNEAFQFRNLPLLEKLIAQLEKADENGGY

Spnp14
MEPHQLTLTVFILSLSVLMAVTSTGAFPQEVRGDRTGHMIDGFSNDIDLLPLQETALIRLLSNLQS
SSSEYASGEDETYPMVASKRSRSGRKLRFCMDVIRNTWRLCRNTRSN

Spnp15
MNTLSQYLLLICSLLVFIQSYALPTYDKQNVDELQGDNDIDEQQLEMWDAMQGGDNDDVFSRLTRG
GEAFSRDRRRVCVSDCSFCHSFFPTYKLGNCFHGCRKGFHDLGCKQFRY

Spnp16
MNLTTCYLAILAAILAVAAGRTLDLGLPVMELQEEDFPQMQEQNMEHQSMRDMVSARLWSIIQRLK
MDQAVDLKDELDTLDQGAEKMLSEDFNKRGRRPARKICINDIWKGRGGGLRCN

Spnp17
MNSTISTLLSLAALLIIAVQMSSALSITEGPQGGSAWALEDNEEPVDYRSVLKLMKYEILLKLMND
LCDELDMCPPSQVPARQAPVVRRGDNNQERRRGGAHLFWRTGVLNKSPIMKAAN

Spnp18
MQPNSIISVAVVMTLATLFTQAVCSLQFETTQDRVPAKRLFWVDKKDHPVDTDFFTVRANDAEEVL
DCFVEVCIADFVNCAKKCLFYENGNTCLPTCRHTRSICSVQCFKRYDVDVSDSVH

Spnp19
MRCYTWVFTVSVFLTSAVLAIASPRWPGGNSQQRPRWELGDADFSSPITDTSFVKRLLGRIHEDLR
QKSNQAADLRDATSRGFETVDLKQLSDNGAGLQVHGVRQTRGKCMGRFGPYMLNCKRSGPTTI

Spnp20
MTSQLVTLVLAVFVCSAAVVYSQSPSSPPSASPPTVLATEPITTPRPAVATTPPPVDNGTPAPSAN
GTDAPTPVVTDAPMTSAKDGDDDGMKGDGDGQKGHDDDEEGGGGLRRGDIALAILATILVVAVICT
FIGLCYWKYKGNSYVTVTADTTYRQ

Figure 3

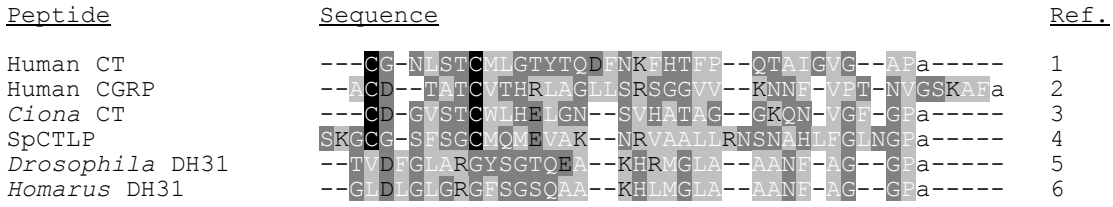
Peptide
Sequence
Ref.
Human CT ---CG-NLSTCMLGTYTQDFNKFHTFP--QTAIGVG--APa----- 1
Human CGRP --ACD--TAICVTHRLAGLLSRSGGVV--KNNF-VPT-NVGSKAFa 2
Ciona CT ---CD-GVSTCWLHELGN--SVHATAG--GKQN-VGF-GPa----- 3
SpCTLP SKGCG-SFSGCMQMEVAK--NRVAALIRNSNAHLFGLNGPa----- 4
Drosophila DH31 --TVDFGLARGYSGTQEA--KHRMGLA--AANF-AG--GPa----- 5
Homarus DH31 --GLDLGLGRGFSGSQAA--KHLMGLA--AANF-AG--GPa----- 6

Figure 4

Peptide	Sequence	Ref.
SpPPLN1d	GFNSGAMEPLGSGFI	1
SpPPLN2h	-FG-GANEPMRSGFF	1
Aplysia PP1-precursor peptide	PLDSVYGTHGMSGFA	2
Aplysia PP2-precursor peptide	PVDSI-GSS----FI	2
Aplysia PP3-precursor peptide	RLDSIAGSSGFSNFG	2
Platynereis FDSIG-precursor peptide	SFDSIGHSSNFAGLD	3
Caenorhabditis NLP14 peptide	AIDGLDGAGF--GFD	4
Caenorhabditis NLP15 peptide	AFDSLAGSGFDNGFN	4
Procambrus orcokinin-precursor peptide	NFDEIDRSGF--GFN	5
Nasonia orcokinin-precursor peptide	NFDEIDRSGF-SGFN	6

Supplementary Material

Click here to download Supplementary Material: Supplementary Figures.doc