

SCIENTIFIC REPORTS



OPEN

Neuropeptide precursors and neuropeptides in the sea cucumber *Apostichopus japonicus*: a genomic, transcriptomic and proteomic analysis

Muyan Chen¹, Alzbeta Talarovicova², Yingqiu Zheng¹, Kenneth B. Storey³ & Maurice R. Elphick^{1,2}

The sea cucumber *Apostichopus japonicus* is a foodstuff with very high economic value in China, Japan and other countries in south-east Asia. It is at the heart of a multibillion-dollar industry and to meet demand for this product, aquaculture methods and facilities have been established. However, there are challenges associated with optimization of reproduction, feeding and growth in non-natural environments. Therefore, we need to learn more about the biology of *A. japonicus*, including processes such as aestivation, evisceration, regeneration and albinism. One of the major classes of molecules that regulate physiology and behaviour in animals are neuropeptides, and a few bioactive peptides have already been identified in *A. japonicus*. To facilitate more comprehensive investigations of neuropeptide function in *A. japonicus*, here we have analysed genomic and transcriptomic sequence data and proteomic data to identify neuropeptide precursors and neuropeptides in this species. We identified 44 transcripts encoding neuropeptide precursors or putative neuropeptide precursors, and in some instances neuropeptides derived from these precursors were confirmed by mass spectrometry. Furthermore, analysis of genomic sequence data enabled identification of the location of neuropeptide precursor genes on genomic scaffolds and linkage groups (chromosomes) and determination of gene structure. Many of the precursors identified contain homologs of neuropeptides that have been identified in other bilaterian animals. Precursors of neuropeptides that have thus far only been identified in echinoderms were identified, including L- and F-type SALMFamides, AN peptides and others. Precursors of several peptides that act as modulators of neuromuscular activity in *A. japonicus* were also identified. The discovery of a large repertoire of neuropeptide precursors and neuropeptides provides a basis for experimental studies that investigate the physiological roles of neuropeptide signaling systems in *A. japonicus*. Looking ahead, some of these neuropeptides may have effects that could be harnessed to enable improvements in the aquaculture of this economically important species.

The use of sea cucumbers (class Holothuroidea; phylum Echinodermata) as a foodstuff, known as trepang or haishen, has been a feature of Chinese culinary culture for hundreds of years¹. Recently, due to overfishing of wild-populations, a multibillion-dollar sea cucumber aquaculture industry developed in the 20th and 21st centuries². The sea cucumber species that is cultured most extensively in China is *Apostichopus japonicus*. However, there are challenges associated with the aquaculture of *A. japonicus* and other edible sea cucumber species. For example, there is a need to develop better methods for induction of spawning and to improve the growth and quality of the edible body wall tissue of sea cucumbers produced in aquaculture facilities³. To accomplish this we need to learn more about the biology of these animals, including genetics, neurophysiology, ecophysiology,

¹The Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao, PR, China. ²School of Biological & Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK. ³Institute of Biochemistry, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada. Muyan Chen and Alzbeta Talarovicova contributed equally. Correspondence and requests for materials should be addressed to M.C. (email: chenmuyan@ouc.edu.cn) or M.R.E. (email: m.r.elphick@qmul.ac.uk)

immunology, epidemiology and nutrition. Furthermore, there are many fascinating biological characteristics of sea cucumbers, that include aestivation, evisceration, regeneration, albinism and autolysis⁴. An important advance in our knowledge of the biology of *A. japonicus* has been the determination of the genome sequencing of this species, with two high-quality data sets reported recently^{5,6}. Genomics has provided important insights into biological processes in this species, including visceral regeneration and aestivation, but there are many other aspects of sea cucumber biology that remain to be investigated.

Important regulators of physiological processes and behaviour in animals are neuropeptide signalling molecules that are synthesized and secreted by neurons; these can exert effects locally, as neurotransmitters or neuromodulators, and/or systemically as hormones^{7–11}. Neuropeptides are derived and cleaved from larger precursor proteins that have several features in common, including an N-terminal signal peptide that targets the protein to the regulated secretory pathway and canonical dibasic or monobasic cleavage sites located N-terminal and/or C-terminal to the neuropeptide sequence(s)^{12,13}. Furthermore, some neuropeptides are subject to post-translational modifications, including conversion of an N-terminal glutamine to pyroglutamate, which is protective against aminopeptidases, and conversion of a C-terminal glycine residue to an amide group, which is protective against aminopeptidases^{14,15}.

Investigation of the phylogenetic distribution of neuropeptides and their cognate G-protein coupled receptors has revealed that the evolutionary origin of at least thirty neuropeptide signalling systems can be traced to the common ancestor of bilaterian animals^{16–18}. However, compared to other well-studied invertebrates such as the insect *Drosophila melanogaster*, the nematode *Caenorhabditis elegans* and the mollusc *Aplysia californica*¹¹, our knowledge of neuropeptide signalling systems in *A. japonicus* and other sea cucumbers is still in its infancy. The first paper to report the identification of neuropeptides in sea cucumbers was published in 1992, with the identification of two neuropeptides, GFSKLYFamide and SGYSVLYFamide, isolated from the sea cucumber *Holothuria glaberrima*¹⁹. These two neuropeptides belong to the SALMFamide family of neuropeptides, which were first discovered in starfish²⁰. Immunocytochemical and pharmacological studies on *H. glaberrima* revealed that GFSKLYFamide is widely expressed in the nervous system and other organs and causes relaxation of *in vitro* preparations of the intestine and longitudinal muscle of the body wall^{21,22}. Furthermore, several other myoactive peptides were identified in extracts of the body wall of *A. japonicus*, including NGIWamide and stichopin^{23,24}. Subsequent studies revealed that NGIWamide, in addition to effects on myoactivity²⁵, also acts as a gonadotropic neuropeptide²⁶ and causes stiffening of the body wall collagenous tissue²⁷. On the other hand, stichopin, a 17-amino acid peptide with a disulphide bridge²³, suppresses the stiffening effect of acetylcholine (ACh) on the body wall dermis²⁷. Furthermore, this effect is consistent with the pattern of expression of stichopin in *A. japonicus*²⁸.

With advances in transcriptome sequencing, it has become feasible to conduct a more comprehensive analysis of the occurrence and diversity of neuropeptide signalling systems in sea cucumbers. Thus, several precursors of the myoactive neuropeptides isolated from *A. japonicus* by Iwakoshi *et al.*²³ and Ohtani *et al.*²⁴ were identified by Elphick²⁹ by analysing transcriptome sequence data obtained using 454 sequencing technology³⁰. Furthermore, informed by the identification of transcripts encoding neuropeptide precursors in the sea urchin (*Strongylocentrotus purpuratus*)³¹, several other neuropeptide precursor transcripts were also identified in *A. japonicus*³². With the development of sequencing technologies, further insights into the diversity of neuropeptide precursors in *A. japonicus* and other sea cucumber species have been obtained recently^{33,34}.

The objective of this study was to perform a detailed analysis of neuropeptides in *A. japonicus* by sequencing the transcriptome of neural tissue (circumoral nerve ring; CNR) and combining analysis of these sequence data with mass spectroscopic analysis of CNR extracts so that the structure of mature neuropeptides could be determined. Furthermore, by analyzing the genome sequence of *A. japonicus*^{5,6}, the exon/intron structure and chromosomal (linkage group) locations of neuropeptide precursor genes have been determined in a sea cucumber species for the first time. Collectively, these data provide an important molecular basis for investigation of the physiological roles of neuropeptides in *A. japonicus* and other sea cucumber species.

Materials and Methods

Animals and sample collection. Adult individuals of the sea cucumber *A. japonicus* (80–120 g body mass) were collected from the coast of Qingdao (Shandong, China) in early May, and acclimated in seawater aquaria at 15 °C for ten days before use. The circumoral nerve ring (CNR) was dissected from randomly selected adults (two males and two females) and used for RNA isolation and transcriptome construction. Another four randomly selected adults (also two males and two females) were sacrificed for peptide/protein isolation and neuropeptide identification using mass spectrometry. CNR tissue was immediately frozen in liquid nitrogen prior to storage at –80 °C until used. All experimental protocols were approved by Ocean University of China.

RNA isolation and transcriptome sequencing. Total RNA was extracted from CNR tissue using an RNeasy mini kit with DNase-treatment (74104, Qiagen, Germany), following the manufacturer's instructions. RNA concentration and quality were determined using an Agilent 2100 bioanalyzer. Total RNA from 3 individuals (two males and one female) were pooled for transcriptome construction. The CNR transcriptome was generated following the manufacturer's standard procedures (Illumina, San Diego, USA). High-quality strand-specific libraries were sequenced on the Illumina HiSeq. 2500 platform and 125-bp paired-end reads were generated. Bases were called using the Illumina CASAVA software. A total of 103,348,164 raw reads were first processed using in-house perl scripts and trimmed by removing adaptor sequences, selecting reads with $\geq 50\%$ of low-quality bases and reads with $\geq 10\%$ uncalled bases (Ns). The downstream analysis was only based on clean data (99,698,992 reads) with the above quality control. Transcriptome *de novo* assembly was accomplished using Trinity (2.1.1)³⁵ with default parameters, with the exception of min_kmer_cov, which was set to 1. The 100,517 unigenes generated (with 20,843 over 1000 bp) were then set up for local BLAST analysis using Bioedit

Kisspeptin-type precursor (AjKPP)

MDKIVFPIILLSLLCGTVFASASLA DTNLKDYEDRLDDARERVLKLIAGLLSDDTYQEQTGEQDEDDLAVNIPILEN
LLAENDGEDVIDADDTAELIFESLSNNGRPIDEKRAGSLDCEASCEDEVERRGRQPNRNNAHYRTLFPKRVQRQTF
STVRNTRKSAVKKNKNSRARPPLLPFGK

Somatostatin-type precursor 1 (AjSSP1 or Ajnp19)

MSWRAGVLLCATVLSWVSLTSAHTWDSDTDFEDEVQEP LLSGVDESMLRSLIIRKFTDRIRNSMKILEDMDL
NTELADPETTYRRQEPWSSQTSGLLSTDKISA KRGGACIGRFVPIILHKCVRK

Somatostatin-type precursor 2 (AjSSP2 or Ajnp16)

MIMSGKTNLVILVFGIVITVAAARS LSLDRDFNGAPQEPNEQNFEDLVNEVPNSPTRTVSQRIKDELMAKLLDLLGK
VGDNAGPYEYIPPMNTRPEFSDDDAARFEENLSKRSKSRYNRWCNLVDIWKGGGNSHRCR

Calcitonin-type precursor 1 (AjCTP1)

MKASLAVPITLCMFCYLLVTVTSVTINRPNAGLELSRQYPELYEKILRQLINEPTQYKRSCSNKFGAGAHMKVANA
VLKQNSRGGQFQKFGSAGPKRSFPYLDLEKR RVGGCGDFSGCASLKAGRDLVRAMLRP SKFGSGGP GK

Calcitonin-type precursor 2 (AjCTP2)

MKASLAVPITLCMFCYLLVTVTSVTINRPNAGLELSRQYPELYEKILRQLINEPTQEKRVGGCGDFSGCASLKAG
RDLVRAMLRP SKFGSGGP GK

Pigment-dispersing factor-type precursor 1a (AjPDFP1a)

MQKILVLIIVSVLVILLGLAAATEALPLSDMTQEEQYTDLDDLVNEEYVKRISDNDFAQLRGPHISQFARNKAFLN
RQRNALEYGQKR DAMSADTMDKRNLSDNDVSQSRAAYMNQMLAYRMMSQLLGEAGR

Pigment-dispersing factor-type precursor 1b (AjPDFP1b)

MQKILVLIIVSVLVILLGLAAATEALPVKRISDNDFAQLRGPHISQFARNKAFLNRQRNALEYGQKR DAMSADTMDK
RNLSDNDVSQSRAAYMNQMLAYRMMSQLLGEAGR

Pedal peptide-type precursor 2 (AjPPLNP2)

MKLIILSTIGFVLAACVLTVLGDDENQNEDEKRFGNSNMDPLVHSLIGKRDAEKRFSSQIMDPLRYSLVS KRSD
KRFGNSNMDPLMYSMIGKRDAEKRFSSQIMDPLRYSLVSKRSAEKRFGNSNMDPLVHSLISGKRDAEEFDEE
KRFGYHPMDPLSNLSMSGKRR

Figure 2. *A. japonicus* neuropeptide precursor proteins that are precursors of neuropeptides belonging to known bilaterian neuropeptide families. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey], C-terminal glycine (G) residues that are putative substrates for amidation are shown in orange and putative monobasic/dibasic cleavage sites are shown in green. Peptides confirmed by mass spectrometry are shown with light yellow highlighting. For peptides with a C-terminal glycine residue (orange) highlighted in yellow, mass spectrometric analysis confirmed that the C-terminal glycine residue is converted to an amide group. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

(2) the presence of putative monobasic or dibasic cleavage sites located N-terminal and/or C-terminal to putative bioactive peptides, with reference to known consensus cleavage motifs^{44–46}, and (3) the presence, in some cases, of a C-terminal glycine residue that is a potential substrate for amidation. However, because these characteristics are not necessarily unique to neuropeptide precursors, some of the proteins identified here using BLAST are classified as putative neuropeptide precursors.

De novo-based identification of candidate neuropeptide precursors in *Apostichopus japonicus*.

The transcriptome sequence data were also analyzed using a novel neuropeptide-prediction tool NpSearch, which uses characteristics of neuropeptide precursors (signal peptide, dibasic cleavage sites) to identify candidate novel neuropeptide precursors (<https://rubygems.org/gems/NpSearch>)⁴⁷. As with some of the proteins identified using BLAST (see above), because these characteristics are not necessarily unique to neuropeptide precursors, some of the proteins identified here using NpSearch are classified as putative neuropeptide precursors.

Analysis of the sequences of neuropeptide precursor transcripts identified in *Apostichopus japonicus*.

The protein sequences of candidate neuropeptide precursors and polypeptide hormone precursors were annotated in color as follows: the N-terminal signal peptide, identified using SignalP v.3.0, was colored blue; putative dibasic or monobasic cleavage sites were colored green; and the putative neuropeptide(s) or peptide hormone(s) derived from the precursor were colored red, with C-terminal glycine residues (when present) shown in orange. Figures combining the color-coded precursor sequences were prepared, with the sequences grouped into four categories (Figs. 1–7), as described below.

The sequences of *A. japonicus* neuropeptides/polypeptide hormones were aligned with homologous peptides from other bilaterian species and/or from other echinoderms. Alignments were generated and edited using ClustalW Multiple alignment with default parameters. GeneDoc (<http://genedoc.software.informer.com>) was used to annotate the alignments and prepare alignment figures, which are collected in Supplementary Fig. 1.

Preparation of *A. japonicus* CNR extracts and identification of neuropeptides using mass spectrometry (MS).

Frozen samples of CNR tissue (~100 mg each) were ground into powder under liquid

Glycoprotein hormone alpha-2-type precursor 1 (AjGPA2P1)
 MLLRKAL**TQLLHI**TLSC**LLW**SCGLQ**GT**TADSWRR**PGCHLL**GYKQ**LV**IPGCRA**EE**IRMNACRGY**CM**TY**SF**PSNVY
 TLF**LSGG**NHVL**TS**HG**SC**TIK**TT**HD**I**H**F**TL**Q**ENNHVYENVIKSAAS**CE**CS**LC**DI**N**

Glycoprotein hormone alpha-2-type precursor 2 (AjGPA2P2) (partial)
 MHICHQ**VFC**NMNGDIQ**VT**FLV**TY**ML**LT**SFYCS**D**ADG**TT**PS**WE**Q**PG**CH**LV**G**FA**K**LI**ELDN**CR**TAT**VP**VNAC**RG**F**C**LS
 Y**S**Y**S**DY**E**RY**I**TS

Glycoprotein hormone beta-5-type precursor (AjGPB5P)
 MDPV**YF**SS**LI**Q**F**V**F**IL**P**ALL**S**Q**A**DSL**D**NHPGASRE**Q**TV**D**SS**Q**F**P**IT**N**PM**N**ALRHGT**GP**YRS**GG**SS**L**P**H**EL**R**IE**P**
 DKIL**K**PS**S**E**L**IDS**G**TS**YL**DD**W**CK**A**Y**T**LR**Q**RI**E**EP**G**CI**S**RI**I**T**N**RL**C**Y**G**Q**C**NS**F**Y**I**PK**Q**S**N**D**Y**N**Q**A**F**Q**S**C**S**F**C**K**P**
 Q**K**V**A**Y**I**T**V**T**L**R**C**PG**Q**DP**P**IK**T**K**R**V**K**RI**K**K**R**C**M**AID**V**S

Bursicon alpha-type precursor (AjBALPP or AjBAP)
 M**I**GE**K**Y**C**Y**S**Y**S**H**V**LL**L**L**I**T**L**I**A**I**A**T**G**Q**I**L**D**DK**P**T**R**H**L**D**H**L**A**L**F**PE**K**T**W**CD**V**K**Q**I**Q**Q**E**V**A**L**P**G**C**H**S**K**L**I**P**N**T**V**C**F**G**
 Q**C**F**S**Y**S**I**P**Q**S**M**P**D**K**A**D**P**H**L**Q**Y**C**E**C**CK**P**I**R**L**S**Q**V**K**V**T**L**T**C**P**R**N**D**V**K**V**M**D**K**L**V**E**V**E**A**C**D**C**Q**K**C**F**E**D**P**P**I**Q**D**G**A**I**I**D**D**
 Y**S**Y**S**Y**Y**E**D**D**I**

Bursicon beta-type precursor (AjBBLPP or AjBBP)
 M**S**F**L**L**L**L**L**V**C**A**A**Q**G**V**L**V**H**S**R**N**E**R**E**R**C**E**L**L**S**K**Q**I**N**A**R**M**E**E**Y**D**S**D**L**G**Q**T**V**H**C**R**G**R**I**T**A**Q**Y**C**E**G**S**C**R**S**K**A**I**P**S**V**T**Y**V**S**G**
 F**Q**R**L**C**R**C**C**F**E**T**D**T**R**T**S**V**W**L**D**H**C**V**T**Q**D**R**R**S**L**S**G**R**Y**F**Q**M**R**E**P**T**S**C**S**C**Q**R**C**

Insulin-like peptide precursor (AjILPP)
 M**A**K**L**G**L**A**P**S**G**G**K**P**W**F**S**T**G**S**T**V**L**F**V**A**G**L**S**F**F**Y**V**A**E**G**S**W**Y**C**S**A**P**E**T**V**R**A**I**C**D**G**C**Y**A**G**G**I**H**T**R**A**F**K**R**S**S**D**I**I**S**L**Y**
 K**D**P**L**K**K**S**N**A**L**N**F**L**L**P**R**S**H**T**P**S**S**L**I**K**R**I**R**S**G**F**I**G**E**C**E**K**N**C**E**I**R**E**M**V**F**Y**C**A**E**K**Q**R**E**Y**A**S**F**F**P**R**F**S**E**T**E**Y**D**T**R**R
 T**I**Q**N**

Figure 3. *A. japonicus* neuropeptide precursor proteins that are precursors of neuropeptides belonging to known bilaterian neuropeptide families. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey] and putative monobasic/dibasic cleavage sites are shown in green. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

nitrogen and homogenized using a Polytron in 1 ml lysis buffer containing 8 M urea, 2 mM EDTA, with 1% protease inhibitor Cocktail Set III added (Calbiochem, Germany). Crude extracts were then sonicated mildly three times on ice using a high intensity ultrasonic processor (Scientz, Ningbo Scientz Biotechnology Co., Ltd., China), and then centrifuged for 10 min (20,000 g, 4 °C). Supernatants containing soluble proteins were collected and a 10 kDa NMWCO ultrafiltration membrane (Merck Millipore, Amicon Ultra-0.5, ultracel-10k, Germany) was used to separate peptides and proteins (5000 g, 4 °C, 45 min). Peptide concentration was quantified using a BCA kit (Beyotime, China) and then 5 µg of peptide solution was acidified to pH 2.0–3.0 with 10% TFA and centrifuged for 10 min (20,000 g, 4 °C). The supernatant was then desalted using a C₁₈ membrane (Millipore, Germany) and lyophilized for LC-MS/MS.

Peptides were dissolved in 0.1% formic acid (Solvent A), directly loaded onto a reverse-phase analytical column (15 cm length, 75 µm i.d.) with a linear gradient from 7% to 25% solvent B (0.1% formic acid in 98% acetonitrile) over 24 min, 25 to 36% solvent B for 8 min, then climbing to 80% in 4 min, and holding at 80% for the last 4 min, all at a constant flow rate of 400 nl/min on an EASY-nLC 1000 UPLC system. The resulting peptides were subjected to nanospray ionization (NSI) followed by tandem mass spectrometry (MS/MS) in a Q Exactive™ Plus hybrid quadrupole-Orbitrap mass spectrometer (ThermoFisher Scientific, USA) coupled online to the UPLC. The electrospray voltage applied was 2.0 kV. The m/z scan range was set up from 350 to 1800 for full scan, and intact peptides were detected in the Orbitrap at a resolution of 70,000. Peptides were selected for MS/MS using normalized collision energy (NCE) setting at 28, ion fragments were detected in the Orbitrap at a resolution of 17,500. A data-dependent procedure that alternated between one MS scan followed by 20 MS/MS scans was applied for the top 20 precursor ions above a threshold ion count of 2E4 in the MS survey scan with 15.0 s dynamic exclusion. Automatic gain control was set at 5E4.

The resulting MS/MS data were analyzed using the Mascot search engine (v.1.5.2.8)⁴⁸ against an *A. japonicus* transcriptome sequence database concatenated with a reverse decoy database. Search parameters were set as follows: no enzyme was used, and variable modifications included methionine oxidation, conversion of glutamine to pyroglutamic acid, deamidation of asparagine and C-terminal amidation. Data are available via ProteomeXchange with identifier PXD013818.

Identification of genes encoding neuropeptide precursors in *A. japonicus*. *A. japonicus* neuropeptide precursor transcript sequences identified by analysis of our CNR transcriptome sequence data or otherwise available in the NCBI GenBank database^{26,49,50}, or identified previously but not deposited in GenBank^{24,29,30,32–34,51,52}, were used to identify genes encoding neuropeptide precursors in the genome of *A. japonicus*. This analysis was performed with NCBI Genome Workbench 2.12.8. (<https://www.ncbi.nlm.nih.gov/tools/gbench/>) using the genome sequence reported by Zhang *et al.*⁵, which is publicly available from GenBank under accession number GCA_002754855.1, and the genome sequence reported by Li *et al.*⁶, which is not publicly available but which was kindly made available to us by the authors. To identify scaffolds containing genes encoding neuropeptide precursors, a BLAST search was performed using all available neuropeptide precursor sequences as queries. Then, each transcript sequence was aligned with its corresponding scaffold(s) to determine

L-type SALMFamide precursor (AjL-SALMFaP)

MPKYQ**I**IVPAV**M**CVL**LAAILARTEA**DGELRILNNRLFELTKELEERLREQQLEADLILTEDGDQEIGM**KKVVSRAW**
SPLV**Q**T**G**IA**F**CG**K**R**T**DGLDRARSQ**T**DQ**R**AK**K****T**RS**R**SM**F**GN**T**AL**P**F**G**K**R**AGYI**P**HAQE**I**WDLQ**DA**ANNLDT**F**EEV**P**
KKRMGFTGNTGILL **KRN**ADDTQE

F-type SALMFamide precursor (AjF-SALMFaP)

MAPPS**I**FL**L**IS**V**S**F**LAVL**H**PS**H**T**E**TL**D**H**T**G**A**IV**K**EIQ**M**EL**P**H**D**ME**T**AQ**L**L**L**Q**G**Q**E**LR**D**L**A**DE**L**AL**D**Y**L**NER**S**DD**D**D
TFS**N**IA**K**RE**P**GR**S**ST**T**NN**K**DT**M**Y**E**PF**V**RR**G**VP**P**Y**V**V**K**V**T**Y**G**K**R**S**D**D**K**R**F**K**S**P**F**M**F**G**K**RE**D**L**N**GL**D**K**R**G**S**P**F**M**F**G**K**
REM**A**Q**P**H**L**SE**K**R**A**R**Y**S**P**F**M**F**G**K**R**D**G**AD**E**DE**N**LE**E**V**K**R**G**G**S**A**L**Y**F**G**K**R**V**P**E**L**A**E**S**D**G**G**S**K**L**Y**F**G**K**R**G**H**R**G**G**Q**F****S**
QF**K**F**G**K**R**E**D**G**A**L**G**M**D**A**N**E**D**DE**M**E**Q**K**F**E**K**K**D**A**I**T**Q**N**K**R**F**K**S**S**F**Y**L**G**K**R**N**Y**V**A**E**N**D**M**E**D**L**Q**D**V

AN peptide precursor (AjANPP)

MKE**V**AG**W**CL**L**AC**Y**FL**V**IV**H**P**V**Q**S**E**A**E**I**P**T**E**D**V**G**DD**W**T**T**LE**K**V**F**SL**T**Q**D**Q**L**K**G**E**S**E**A**L**V**K**N**D**K**S**D**E**E**V**E**E**T**E**D**E
KR**A**N**R**Y**N**A**L**R**G**R**Y**P**K**Y**R**V**S**G**R**S**D**N**T**D**A**I**D**E**V**R**N**L**Q**D**Q**M**E**D**E****K**R**A**N**R**R**F**S**V**G**K**S**R**F**R**L**R**T**R**S**E**N**D**E**N**L**D**D**E**L**E**E**D**
KR**A**N**R**Y**N**A**L**R**E**E**K**R**G**N**R**H**S**L**R**G**S**A**R**N**R**Y**R**L**R**T**R**N**G**D**G**E**E**E**E**E**E**D**G**E**E**E**E**E**E**N**N**I**K**Q**E**L**A**N**E**G**D**E**E**E**G**D**P**T**D**P**E**
KR**A**N**R**Y**N**A**L**R**G**K**S**R**Y**R**I**R**T**R**S**D**E**T**W**L**N**D**E**E**Q**Q**L**E**E**K**R**G**N**R**H**S**L**R**G**S**A**R**N**R**Y**R**L**R**T**R**N**G**D**E**D**E**E**E**E**E**D**G**E**E**E**
EE**N**I**K**Q**E**L**A**N**E**G**T**R**K**K**E**I**H**S**R**E**K**S**E**Q**I**N**A**L**R**G**K**Q**D**T**G**S**E**Q**D**Q**M**M**L**S**S**M**M**K**N**I**R**R**E**E**G**N**R**Q**S**S**L**R**G**S**A**R**N**R**Y**
RL**R**T**R**N**G**D**G**E**E**E**T**W**L**N**D**D**E**D**Q**L**E**E**K**R**G**N**R**Q**S**S**L**R**G**S**A**R**N**R**Y**R**L**R**T**R**N**G**D**E**D**E**E**E**E**E**E**E**E**K**E**E**V**K**E**E**E**
EE**E**E**D**P**T**D**P**E**K**R**A**N**R**Y**N**A**L**R**G**K**S**R**Y**R**I**R**T**R**S**D**E**T**W**L**N**D**D**E**Q**Q**L**E**D**K**R**G**N**R**Q**S**S**L**R**G**S**A**R**N**R**Y**R**L**R**T**R**N**G**D**E**
EE**E**K**K**K**T**R**R**R**R**R**S**R**I**G**G**R**S**Q

Neuropeptide precursor 9 (Ajnp9)

MV**K**L**K**I**R**C**V**F**L**P**L**C**K**L**T**S**V**A**V**L**K**L**L**P**H**G**R**S**F**G**K**R**S**S**R**D**R**S**Q**L**K**L**I**L**G**V

Neuropeptide precursor 11 (Ajnp11)

MN**S**F**L**V**C**L**L**S**A**I**L**L**A**G**V**L**G**K**P**T**F**D**F**D**D**L**I**E**K**R**G**A**D**R**A**L**T**T**R**C**L**K**D**C**M**F**C**A**K**Y**T**K**N**G**F**S**I**R**Q**C**M**Q**E**C**G**G**E**V**A**S**G**R
EQ**T**W**A**S**C**D**M**F**S**N**G**K

Figure 4. *A. japonicus* proteins that are precursors of neuropeptides (SALMFamides and AN peptides) and putative neuropeptides (Ajnp8, 9 and 11) that have thus far been found only in echinoderms. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey], C-terminal glycine (G) residues that are putative substrates for amidation are shown in orange and putative monobasic/dibasic cleavage sites are shown in green. Peptides confirmed by mass spectrometry are shown with light yellow highlighting. For peptides with a C-terminal glycine residue (orange) highlighted in yellow, mass spectrometric analysis confirmed that the C-terminal glycine residue is converted to an amide group. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

Neuropeptide precursor 15 (Ajnp15)

MS**Q**L**S**S**V**C**L**L**L**V**L**L**S**V**T**V**Y**A**Y**P**A**D**Q**A**G**N**L**L**D**L**D**E**L**F**Q**A**Q**E**I**P**V**D**A**P**E**N**M**P**F**G**R**E**Q**K**R**L**G**S**A**T**C**R**R**E**C**A**Y**C**A**K**I**H**T**R
YS**L**G**R**C**F**R**T**C**M**T**G**K**T**D**Y**S**C**A**N**R**M**H**F**V

Neuropeptide precursor 18 (Ajnp18)

MN**V**S**H**L**M**Y**I**A**I**F**A**V**S**A**F**M**I**H**A**V**S**S**L**S**Y**N**T**I**D**D**T**I**P**Y**K**R**L**F**W**I**D**K**K**A**S**Y**D**S**P**L**Y**T**V**R**S**S**A**P**I**I**V**D**C**F**L**K**E**C**I**K**S**
FS**A**C**G**K**T**C**I**M**T**N**E**E**M**P**C**M**A**A**C**N**Q**Q**Q**L**C**S**T**R**C**F**Q**R**Y**A**V**K**T**V**G**N**A**A

Neuropeptide precursor 23 (Ajnp23)

MN**L**T**V**K**L**L**L**L**A**V**A**I**C**L**F**I**I**C**A**V**E**T**A**E**Q**V**N**D**S**S**D**L**D**G**D**Q**E**G**Y**E**D**E**Y**P**L**L**N**L**L**D**K**L**V**A**S**R**Q**R**D**T**R**N**W**R**S**N**P**K**F**R**N**Y**K**A
HS**A**G**W**K**R**A**P**L**A**D**D**T**A**H**Q**V**D**E**R**R**Y**N**S**G**R**Y**S**G**Y**G**Q**G**N**R**C**S**A**G**W**K**R**T**S**K**T**Q**D**L**S**

Neuropeptide precursor 25 (Ajnp25)

MS**F**D**S**V**S**K**I**T**M**K**V**I**M**T**F**L**L**C**L**V**M**L**Q**Y**V**C**H**A**K**S**A**C**S**N**R**H**P**K**L**C**I**L**H**P**G**K**R**G**Y**I**S**A**A**S**N**D**R**P**G**N**E**N**G**V**S**I**D**R**T**E**F**T**Q**
PK**W**L**H**D**E**L**T**K**S**G**E**L**E**N**L**F**G**Q**E**E**D**I**I**P**F**N**D**D**R**L**G**Y**T**Q**D**L**L**T**K**V**G**I**D**H**L**E**R**R**E**Q**I**L**R**I**L**T**S**L**L**E**D**D**L**S**Y**R

Figure 5. *A. japonicus* proteins that are precursors of putative neuropeptides that have thus far been found only in echinoderms. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey], C-terminal glycine (G) residues that are putative substrates for amidation are shown in orange and putative dibasic cleavage sites are shown in green. Peptides confirmed by mass spectrometry are shown with light yellow highlighting. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

the exon-intron structure of the gene using SPLIGN/ProSPLIGN alignment tools integrated in NCBI Genome Workbench^{53,54} and to determine its position in the genome with reference to scaffold maps of twenty-two linkage groups (chromosomes). The GCA_002754855.1 genome⁵ was used to determine the location of exons and introns in neuropeptide precursor genes and the genome sequence data obtained by Li *et al.*⁶ was used to determine the location of neuropeptide precursor genes on linkage groups. Only alignments with at least 90% identity were further examined, with some exceptions highlighted in yellow in Supplementary File 1. Spreadsheets containing details of the data obtained, including protein sequences and their length, available transcript GenBank IDs,

GLRFA precursor (AjGLRFA-P)

MKLLTVLLLSCLLMSLVLTTDARRHPCKRGLRFARHPCPKPREAEADDHGAFQRRREENAEAD

GN19 precursor (AjGN19P)

MYKTAEKFFLICAVLTTALLLVCNSMYIQRDDAPDGLDRYLEMKRGQMMMRMDLLEESLKRGGRL
PNYAGPPRMPWLIHNRR

Stichopin precursor (AjStichopinP)

MFARAVILLAVLACAALTTVSADRQGWPACYDSKGNKYC

SWYG precursor Aj_SWYGP2

MALIRCLLAFAFIVCSVKTFAAP IEDEILDADALSRRGGWFHRLFHHHHHGAPADNKNQPKQREE
 RQLIHDNRENAMLDDFEKKMAELSPAERFDWHALLHSVSKHVAKHVAENLKARVLAEEGERLPEG
 YDELTPEERSWYGTLGKAALGALSFLGK

SWYG precursor Aj_SWYGP3

MALIRCLLAFAFIVCSVKTFAAP IEDEILDADALSRRGGWFHRLFHHHHHGAPADNKNQPKQREE
 RQLIHDNRENAMLDDFEKKMAELSPAERFDWHALLHSVSKHVAKHVAENLKARVLAEEGERLPEG
 YDELTPEERSWYGSLASALLR

APHAIRPPSG precursor

MNLKVMVLVLVSMILFINAAPHAIRPPSGRSACDCGCAFTGGVGCAPEDCIRG

GYWKDLNDNYVKAHKT precursor

MSLLKVIALVLAFTTVTTLASPSNLKDASEADFLSELFKRDDGFDEKALIPEDRGYWKDLNDNYVK
AHKTRGISDDQGLPDALSERLAELSLDDDFGSKVFDVMARDLKMKGLOQYEGEDLPDRFEDLTPLQ
 RRRWRFSYKMAFRVVGKILRALLPLIRVLG

MPMNPADYFSRGTVYIPTRDS precursor

MHFQVLTRTLLEFVCLCFIAAVTSMPMNPADYFSRGTVYIPTRDSRSSVPVEPGMMGHMCVYHP
 NFYCMCTSNGFISFCENHSSFFSKK

Figure 6. Neuropeptide precursors that are precursors of bioactive peptides have been discovered previously in *A. japonicus*. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey], C-terminal glycine (G) residues that are putative substrates for amidation are shown in orange and putative monobasic/dibasic cleavage sites are shown in green. Peptides confirmed by mass spectrometry are shown with light yellow highlighting. For peptides with a C-terminal glycine residue (orange) highlighted in yellow, mass spectrometric analysis confirmed that the C-terminal glycine residue is converted to an amide group. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

number of exons, scaffold locations, GenBank IDs of predicted proteins, linkage group, references, and details of SPLIGN/ProSPLIGN^{53,54} alignments for each genome analysis are provided in Supplementary File 1.

Results and Discussion

By analyzing novel *A. japonicus* circumoral nerve ring (CNR) transcriptome sequence data in combination with other publicly available *A. japonicus* transcriptome sequence data, we have identified 44 candidate neuropeptide precursors. For the purposes of discussion we have divided these into four groups: (1) Precursors of neuropeptides belonging to known bilaterian neuropeptide families (Figs. 1–3), (2) Precursors of neuropeptides belonging to neuropeptide families that have thus far only been identified in echinoderms and precursors of other putative echinoderm neuropeptides (Figs. 4–5), (3) Precursors of neuropeptides that have been identified as myoactive neuropeptides in *A. japonicus* (Fig. 6), and (4) Other novel putative neuropeptide precursors identified in *A. japonicus* based on their sequence similarity with known *A. japonicus* precursor proteins (Fig. 7). The sequences of these proteins are shown in Figs. 1–7, with the predicted N-terminal signal peptide shown in blue, predicted neuropeptides derived from each protein shown in red, monobasic or dibasic cleavage sites shown in green, C-terminal glycine residues that are potential substrates for amidation shown in orange and cysteine residues highlighted grey. Where the structures of the predicted neuropeptides were confirmed by mass spectroscopic analysis of CNR extracts (see Table 1 and Supplementary Figs. 2 and 3), the neuropeptide sequence is highlighted in yellow. Amino acids or pairs of amino acids whose codon or codons are interrupted by an intron in the gene encoding the neuropeptide precursor are underlined.

***A. japonicus* proteins that are precursors of neuropeptides belonging to known bilaterian neuropeptide families.** *Vasopressin/oxytocin-type precursor* (“holotocin” precursor). An *A. japonicus* vasopressin/oxytocin-type neuropeptide precursor (or holotocin precursor) was identified that was consistent with a previously reported sequence³⁴. It is a 163 amino acid protein with a predicted 24-residue N-terminal

Peptide name	Sequence	Modifications	m/z	Charge	Predicted avg mass (Da)	Precursor length AA
AjTRH	QLPGSPWKFE	Gln- >pyro-Glu,acylamide	678.840405	2	1355.66626	454
AjKP	GRQPNRNAHYRTLPP	Acylamide	609.327549	3	1824.96082	180
AjPPLN	FGNSNMDDLPLVHSLIG	Oxidation (M),acylamide	808.398489	2	1614.78243	220
	FGSSQIMDPLRYSLSV	Oxidation (M),acylamide	907.958911	2	1813.90327	
	FGNSNMDDLPLMYSMIG	Acylamide	838.367486	2	1674.72042	
AjL-SALMFa	RMGFTGNTGILL	Acylamide	639.85299	2	1277.69143	178
AjANP	ANRYNALR	Unmodified	489.267407	2	976.520262	555
Ajn23	APLADDTAHQVDE	Unmodified	691.3	2	1380.62	129
AjGN19	GQMMMRDMDLLEESL	Unmodified	899.89575	2	1797.77695	82
	GGRLPNYAGPPRMPWLIHN	Unmodified	716.04242	3	2145.10543	
AjSWYG-3	SWYGLASALL	Acylamide	583.813856	2	1165.61316	153
AjAPHAIRPPSG	APHAIRPPSG	Unmodified	473.266876	2	944.519199	56
AjMPMNPADYFSRGTVYIPTRDS	MPMNPADYFSRGTVYIPTRDS	Oxidation (M)	812.043715	3	2433.10932	91

Table 1. List of neuropeptide precursor-derived peptides detected by mass spectrometry in *A. japonicus* CNR extract. m/z, mass to charge.

signal peptide, a VP/OT-type neuropeptide sequence (CFITNCPLGG) followed by a dibasic cleavage site and a C-terminal neurophysin domain (Fig. 1, GenBank accession number MF401997). Based upon the known structures of VP/OT-type neuropeptides^{55,56}, the predicted structure of the mature VP/OT-type neuropeptide (holotocin) derived from this precursor is CFITNCPLG-NH₂, with a disulfide bridge between the two cysteine residues (underlined) and a C-terminal amide group (see Supplementary Fig. 1A for an alignment of holotocin with VP/OT-type peptides from other taxa).

VP/OT-type neuropeptides have been identified in many vertebrate^{57,58} and invertebrate^{59–62} species and have a variety of physiological roles, which include regulation of reproductive behavior and associative learning⁶³. However, very little is known about the physiological roles of VP/OT-type neuropeptides in echinoderms. The VP/OT-type neuropeptide echinotocin has been found to act as myostimulatory peptide in the sea urchin *Echinus esculentus*, causing contraction of tube foot and oesophagus preparations³⁸. Furthermore, analysis of the expression of the precursor of the VP/OT-type neuropeptide asterotocin in larvae of the starfish *Asterias rubens* revealed expression in the attachment complex, suggesting a potential role for asterotocin in the attachment process prior to metamorphosis⁶⁴. With the discovery of the holotocin precursor in *A. japonicus*, an opportunity to investigate the physiological roles of a VP/OT-type neuropeptide in sea cucumbers has been provided.

NPS/CCAP-type precursor (NGIYWamide precursor). The neuropeptide NGIYWamide was originally discovered as a myoactive neuropeptide in *A. japonicus*^{23,24}. Subsequently studies have revealed that NGIYWamide also causes stiffening of body wall connective tissue²⁷ and stimulation of gamete release in *A. japonicus*²⁶. Identification of the receptor for a related neuropeptide in sea urchins (NGFFFamide) has revealed that NGIYWamide belongs to a bilaterian family of neuropeptides that includes neuropeptide-S in vertebrates and crustacean cardioactive peptide (CCAP) in protostomian invertebrates⁶⁵. An alignment that compares the sequence of NGIYWamide with other “NG peptides” that have been identified in invertebrate deuterostomes and with human neuropeptide-S is shown in Supplementary Fig. 1B. The sequence of the NGIYWamide precursor identified here (Fig. 1, GenBank accession number MF401992) from analysis of CNR transcriptome data is identical to a previously reported sequence²⁹.

Gonadotropin-releasing hormone-type precursor (AjGnRHP) and corazonin-type precursor (AjCRZP). Analysis of CNR transcriptome sequence data enabled identification of the partial sequence of an *A. japonicus* GnRH-type neuropeptide precursor (AjGnRHP), which is the first GnRH-type precursor to be discovered in a sea cucumber species. AjGnRHP comprises a predicted 19-residue N-terminal signal peptide and a putative GnRH-type peptide sequence (QLLGNVQLPLPGG) followed by a dibasic cleavage site (Fig. 1, GenBank No. MF401983). The presence of an N-terminal glutamine residue and a C-terminal glycine residue are consistent with GnRH-type neuropeptides that have been identified in other species⁶⁶ and are indicative of post-translational modifications giving rise to an N-terminal pyroglutamate residue and a C-terminal amide group in the putative mature peptide.

Analysis of CNR transcriptome sequence data also enabled identification of a GnRH-related corazonin-type precursor (AjCRZP), the sequence of which has been reported previously³⁴. AjCRZP is a 111-residue protein comprising a predicted 26-residue N-terminal signal peptide and a corazonin-type peptide sequence (HNTYSMKGKYRWRAG) followed by a dibasic cleavage site (Fig. 1, GenBank accession number MF401982). The presence of a C-terminal glycine residue is indicative of a probable post-translational modification giving rise to a C-terminal amide group in the putative mature peptide.

An alignment of AjGnRH and AjCRZ with GnRH/CRZ-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1C. By comparison with the GnRH-type peptides that have been identified in other echinoderms, the sequence of AjGnRH is atypical. For example, it lacks a tryptophan residue in the C-terminal region, which is a conserved feature of many GnRH-type peptides in other taxa. However, the

A. japonicus GLRFA precursor-like precursor (AjGLRFALP)

MKLVVYVLMLSFLVVSIVATTVA**RHP****CPDKRNVHRDHP****CPD****KKR**TFSLRDQLFQQREENADAD

A. japonicus neuropeptide precursor 11-like precursor (Ajnp11LP)

MRTLIFSCGLMLLGTIAAEMEYDREDLMNFISSLLQSGTGESAGPAVGDENIDTDGDNELAYPM
V**GR****G****K****M****T****Q****E****I****C****R****R****E****C****S****S****C****S****K****V****M****N****V****L****F****M****S****C****W****R****D****C****Q****L****S****S****T****G****R****S****P****H****N****V****W****S****T****C****R****R****F****L****T****G**

A. japonicus neuropeptide precursor 14-like precursor (Ajnp14LP)

MAFVREIMLLLIIVLRRS**WR****TP****T****W****R****T****C****S****V****G****C****K****C****R****W****C****M****D****M****V****R****I****S****W****C****S****V****E****I**

Figure 7. Novel putative neuropeptide precursors in *A. japonicus* based on NP search analysis or BLAST analysis using known *A. japonicus* neuropeptide precursors as queries. Predicted signal peptides are shown in blue, putative neuropeptides are shown in red [with cysteine (C) residues highlighted in grey], C-terminal glycine (G) residues that are putative substrates for amidation are shown in orange and putative monobasic/dibasic cleavage sites are shown in green. The positions of introns in the open reading frame of the gene encoding each neuropeptide precursor are indicated by underlining the amino acid residue(s) whose codon(s) are interrupted by introns.

position of the AjGnRH peptide in the precursor protein immediately after the signal peptide, and the presence of a predicted C-terminal PG-NH₂ motif and a predicted N-terminal pyroglutamate are features that indicate that it is an ortholog of GnRH-type peptides that have been characterised in other taxa. This issue could be investigated further by identification of the receptor for AjGnRH, as has been accomplished in other echinoderms⁶⁷. In contrast to AjGnRH, AjCRZ shares higher sequence similarity with the CRZ-type peptides that have been identified in other echinoderms (Supplementary Fig. 1C), including the peptide ArCRZ that has been shown to act as a ligand for a corazonin-type receptor in the starfish *A. rubens*⁶⁷.

GnRH-related neuropeptides have been identified in a variety of vertebrate⁶⁸ and invertebrate^{66,69–72} species. Corazonin (CRZ) is a homolog of GnRH that was first discovered in cockroaches⁷³ and subsequently found to have a variety of roles in insects, including triggering initiation of ecdysis in moths, triggering gregarization-associated dark-pigmentation in locusts and modulating stress and metabolism^{74–76}. Recently, a neuropeptide (ArCRZ) that acts as a ligand for a corazonin-type receptor in the starfish *A. rubens* was identified⁶⁷ and on this basis the existence of corazonin-type neuropeptides in echinoderms was first established. Accordingly, AjCRZ shares sequence similarity with ArCRZ and with a CRZ-type neuropeptide (Spn12 or SpCRZ) in the sea urchin *S. purpuratus*^{31,34} (Supplementary Fig. 1C).

Nothing is known about the physiological roles of GnRH-type and CRZ-type neuropeptides in sea cucumbers. However, it was recently demonstrated that both ArGnRH and ArCRZ are myoexcitatory neuropeptides in starfish, causing contraction of cardiac stomach, apical muscle, and tube foot preparations, but with different potencies⁷⁷. The discovery of both AjGnRHP and AjCRZP, as reported here, provides a basis for investigation and comparison of the physiological roles of GnRH-type and CRZ-type neuropeptides in sea cucumbers.

Cholecystokinin-type precursors AjCCKP1 and AjCCKP2. The *A. japonicus* cholecystokinin-type precursor 1 (AjCCKP1) was identified here for the first time and named according to its similarity to the *S. purpuratus* CCK-type precursor 1 (SpCCKP1)³³. AjCCKP1 is a 163-residue protein comprising a predicted 24-residue N-terminal signal peptide followed by a 139-residue sequence (residues 25–163) that contains putative dibasic cleavage sites (KR) at residues 109/110, 121/122 and 146/147, and a monobasic cleavage site (R) at residue 135 (Fig. 1, GenBank accession number MH636358). Putative neuropeptides derived from this precursor are AjCCK1.1 (DYGD LGFFF-NH₂), a predicted C-terminally amidated 9-residue peptide formed by residues 110–120 and AjCCK1.2 (DYND LGMFF-NH₂), a predicted C-terminally amidated 9-residue peptide formed by residues 136–145. AjCCK1.1 and AjCCK1.2 share sequence similarity and have in common the motif DYxDLGx⁺FF-NH₂ (where x = variable). The underlined tyrosine residue is predicted to be sulphated, based on the occurrence of this post-translational modification in CCK-type neuropeptides in vertebrates and other invertebrates⁷⁸.

The *A. japonicus* cholecystokinin-type precursor 2 (AjCCKP2) identified here has been reported previously³³ and is a 175 amino acid protein with a predicted 25-residue N-terminal signal peptide and two putative cholecystokinin-type peptides bounded by mono- or dibasic cleavage sites: AjCCK2.1, MNGWYTG⁺GMF-NH₂, and AjCCK2.2, NIPQTYLSGDYF-NH₂ (Fig. 1, GenBank accession number MH351773). As with the peptides derived from AjCCKP1, the underlined tyrosine residues are predicted to be sulphated, based on the occurrence of this post-translational modification in CCK-type neuropeptides in vertebrates and other invertebrates⁷⁸.

An alignment of the *A. japonicus* CCK-type neuropeptides with CCK-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1D. A conserved feature of many CCK-type peptides in other taxa is a Tyr-Gly (YG) motif and this is also present in AjCCK1.1, but because of the presence of an additional amino acid in AjCCK1.1 the YG motif in this peptide does not align with the YG motifs of CCK-type peptides in other taxa (Supplementary Fig. 1D). Further studies are now needed to determine the identity of receptors that mediate the effects of the peptides derived from AjCCKP1 and AjCCKP2 in *A. japonicus* as this may provide insights into their relationships with CCK-type peptides that have been characterised in other taxa.

CCK-type neuropeptides are involved in regulation of a variety of physiological processes, but they are perhaps best known as regulators of food intake in both vertebrates and insects⁷⁸. Consistent with this role, antibodies

to mammalian CCK reveal immunolabelled cells and processes in the intestine of sea cucumbers and synthetic mammalian CCK-type peptides were found to cause relaxation of sea cucumber intestinal muscle⁷⁹. Now with the discovery of the sequences of CCK-type peptides in sea cucumbers and other echinoderms, further investigation of the physiological roles of CCK-type neuropeptides in these animals is feasible.

Thyrotropin-releasing hormone (TRH)-type precursor (AjTRHP). The *A. japonicus* thyrotropin-releasing hormone-type precursor (AjTRHP) identified here has been reported previously³² and is a 453 amino acid protein with a predicted 29-residue N-terminal signal peptide and 19 putative TRH-type peptides bounded by mono- or dibasic cleavage sites: AjTRH1 (pQYFA-NH₂; 10 copies), AjTRH2 (pQLPG-NH₂; 1 copy), AjTRH3 (pQFFQ-NH₂; 4 copies), AjTRH4 (pQHFV-NH₂; 1 copy), AjTRH5 (pQHFA-NH₂; 1 copy), AjTRH6 (pQHFL-NH₂; 1 copy), and AjTRH7 (pQLPGSPWKFE-NH₂; 1 copy) (Fig. 1, GenBank accession number MF401984). Our MS analysis did not detect the tetrapeptides because only peptides longer than seven amino acids or shorter than twenty-three amino acids can be detected with the methods employed. However, AjTRH7 was detected, confirming the existence of post-translational modifications where the N-terminal glutamine (Q) is converted to pyroglutamate (pQ) and a C-terminal glycine is converted to an amide group (Fig. 1, Supplementary Figs. 2A and 3A). These post-translational modifications are also characteristics of TRH in vertebrates⁸⁰ and therefore it is likely that the same post-translational modifications occur in the TRH-like tetrapeptides in *A. japonicus*. An alignment of the *A. japonicus* TRH-type neuropeptides with TRH-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1e.

The physiological roles of TRH in mammals and other vertebrates are well characterised⁸¹ and recently analysis of TRH-type signalling in the nematode *C. elegans* revealed evidence of an evolutionarily conserved role in regulation of growth⁸². Currently, little is known about the physiological roles of TRH-type signalling in echinoderms. However, analysis of the expression of the TRH-type precursor in larvae of the starfish *A. rubens* revealed a potential role for TRH-type peptides in the attachment process prior to metamorphosis⁶⁴. Investigation of the functions of TRH-type peptides in both larval and adult echinoderms, including sea cucumbers, may provide key insights into the evolution of the physiological roles of this neuropeptide family.

Orexin-type precursors (AjOXP1 and AjOXP2). The *A. japonicus* orexin-type precursors (AjOXP1 and AjOXP2) identified here have been reported previously³⁴. AjOXP1 is a 152-residue protein comprising a predicted 31-residue N-terminal signal peptide and an OX-type peptide sequence (DRRCCEQVQGCRIPRNCRCFVKEHVCRQSARNKFTLG) followed by a dibasic cleavage site (Fig. 1, GenBank accession number MF401988). As for OX-type neuropeptides identified in other echinoderms (*A. rubens* and *S. purpuratus*), the presence of a C-terminal glycine residue is indicative of a post-translational modification that gives rise to a C-terminal amide group on the mature peptide. AjOXP1 also contains six cysteine residues, which may form up to three disulfide bridges. A second OX-type neuropeptide precursor in *A. japonicus* (AjOXP2) is a 138-residue protein comprising a predicted 26-residue N-terminal signal peptide and an OX-type peptide sequence, followed by a dibasic cleavage site (Fig. 1, GenBank accession number MF401989). As with AjOXP1, the presence of a C-terminal glycine residue suggests a post-translational modification to give a C-terminal amide group and the presence of six cysteine residues is indicative of three disulfide bridges in the mature peptide.

All of the OX-type peptides identified in echinoderms, including *A. japonicus*, are predicted to contain six cysteine residues, consistent with the OX-type peptide in the hemichordate *S. kowalevskii*^{16,17}. Thus, this feature is a common character shared between hemichordates and echinoderms (collectively ambulacrarians) (see alignment in Supplementary Fig. 1F). OX-type neuropeptides have been reported to be involved in regulating feeding behavior and the period of sleep versus wakefulness in mammals and teleost fish^{83–86}, but nothing is known about the physiological roles of OX-type peptides in echinoderms. The identification of OX-type neuropeptides in *A. japonicus* provides opportunities to address this issue in a sea cucumber species.

Luqin-type precursor (AjLQP). The *A. japonicus* luqin-type precursor (AjLQP) identified here has been reported previously³² and is a 115-residue protein comprising a predicted 30-residue N-terminal signal peptide and a luqin-type peptide sequence (KPYKFMRWG) followed by a dibasic cleavage site (Fig. 1, GenBank accession number MF401981). The predicted luqin-type neuropeptide derived from AjLQP is KPYKFMRW-NH₂, with conversion of a C-terminal glycine to an amide group.

Analysis of the phylogenetic distribution of luqin-type neuropeptide signaling has revealed that it can be traced back to the common ancestor of protostomes and deuterostomes but with loss in the chordate lineage^{16,17,87}, and an alignment of AjLQ with luqin-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1G. The physiological roles of luqin signalling are well-characterised in protostomes⁴² and recently the first insights into the physiological roles of luqin signalling in an echinoderm were obtained from experimental studies on the starfish *A. rubens*⁸⁷. Two receptors for the *A. rubens* luqin-type peptide ArLQ were identified and the expression pattern of the ArLQ precursor transcript was mapped. It was discovered that ArLQ acts to cause relaxation of tube feet in *A. rubens*. Further investigation of luqin-type neuropeptide function in echinoderms is now needed, including the AjLQ peptide reported here and previously.

Kisspeptin-type precursor (AjKPP). The *A. japonicus* kisspeptin-type precursor (AjKPP) identified here has been reported previously³⁴ and is a 180-residue protein comprising a predicted 23-residue N-terminal signal peptide and two putative KP-type peptides: AjKP1 and AjKP2, which both have a predicted C-terminal LxF-NH₂ motif (Fig. 2, GenBank accession number MF401998). Furthermore, our mass spectrometric analysis of CNR extracts confirmed the predicted structure of AjKP1 as the C-terminally amidated peptide GRQPNRNAHYRTLPP-NH₂ (Supplementary Figs. 2B and 3B). This is important because, in addition to the dibasic cleavage site (RR) that precedes the ArKP1 peptide sequence in AjKPP, there is another dibasic cleavage site (KR) located at positions

110 and 111 in AjKPP (Fig. 2) and if cleavage did not occur at the RR site a much larger peptide containing two cysteine residues in its N-terminal region would be generated. Previous studies have predicted that a longer peptide of this type is derived from KP-type precursors in other echinoderms (e.g. ArKP1, 42; SpKP1 and OvKP1, 33). Therefore, our finding that the shorter peptide GRQPNRNAHYRTLPPF-NH₂ (ArKP1) is derived from AjKPP in *A. japonicus* is important and may likewise be applicable to other echinoderms.

We conclude that two neuropeptides are derived from AjKPP, AjKP1 (GRQPNRNAHYRTLPPF-NH₂) and AjKP2 (SAVKNKNKSRRARPLLPPF-NH₂), and an alignment of AjKP1 and AjKP2 with kisspeptin-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1H. Further insights into the relationships of AjKP1 and AjKP2 with kisspeptin-type peptides that have been characterised in vertebrates could be achieved by identification of the receptors that mediate effects of AjKP1 and/or AjKP2 in *A. japonicus*.

Kisspeptin-type neuropeptides are well characterised as regulators of reproductive maturation in vertebrates⁸⁸. However, little is known about the physiological roles of kisspeptin-type neuropeptides in invertebrates. The discovery and structural characterisation of AjKP1 and AjKP2, as reported here, provides a basis for experimental studies on these neuropeptides in sea cucumbers.

Somatostatin-type precursors (AjSSP1 or Ajnp19 and AjSSP2 or Ajnp16). The *A. japonicus* somatostatin-type precursors (AjSSP1 and AjSSP2) identified here have been reported previously, but were originally referred to as Ajnp19 and Ajnp16³² and then subsequently identified as somatostatin-type precursors³³. AjSSP1 is a 129 residue protein, comprising a predicted 24-residue N-terminal signal peptide and a predicted 19-residue somatostatin-type peptide (AjSS1) located in the C-terminal region of the precursor (Fig. 2, GenBank accession number MF401987). AjSSP2 is a 139-residue protein comprising a predicted 23-residue N-terminal signal peptide and a predicted 27-residue somatostatin-type peptide (AjSS2), preceded by a dibasic cleavage site (Fig. 2, GenBank accession number MF402007). Both AjSS1 and AjSS2 contain two cysteine residues, which are predicted to form a disulfide bridge based on the occurrence of this post-translational modification in somatostatin-type peptides that have been structurally characterised in other taxa⁸⁹ (see alignment in Supplementary Fig. 1I). Discovery of the receptors that mediate effects of AjSS1 and/or AjSS2 in *A. japonicus* would provide further insights into their relationships with the SS-type peptides that have been characterised in vertebrates.

The physiological roles of somatostatin-type peptides are well characterised in vertebrates⁹⁰ but nothing is known about the functions of somatostatin-type peptides in echinoderms. Identification of AjSS1 and AjSS2 will facilitate investigation of the physiological roles of somatostatin-type peptides in sea cucumbers.

Calcitonin-type precursors (AjCTP1 and AjCTP2). The *A. japonicus* calcitonin-type precursor (AjCTP1) identified here has been reported previously³². AjCTP1 is a 147-residue protein comprising a predicted 23-residue N-terminal signal peptide and two putative CT-type peptides bounded by dibasic cleavage sites: AjCT1 and AjCT2 (Fig. 2, GenBank accession number MF401985). The presence of a C-terminal glycine residue and two cysteine residues in the N-terminal region of these peptides are consistent with post-translational modifications that occur in other calcitonin-type peptides – C-terminal amidation and a disulfide bridge, respectively. A transcript encoding a second calcitonin-type precursor was also identified in *A. japonicus* - AjCTP2 (Fig. 2, GenBank accession number MF401986), which is a shorter form of ArCTP1 that comprises the same signal peptide as in AjCTP1 as well as AjCT2, but not AjCT1. The existence of these long and short calcitonin precursor isoforms in *A. japonicus* has been reported previously and attributed to alternative splicing of transcripts derived from the same gene³⁴. An alignment of AjCT1 and AjCT2 with calcitonin-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1J. This highlights the conserved cysteine residues in the N-terminal region and the C-terminal Pro-NH₂ motif, which are conserved features of CT-type peptides.

Calcitonin-related neuropeptides are involved in regulation of a variety of physiological processes in mammals, including acting as potent and powerful vasodilators and causing relaxation of intestinal longitudinal muscle and inhibition of intestinal peristalsis⁹¹. Furthermore, calcitonin-related peptides that act as diuretic hormones have been identified in insects⁹². Recently, the expression of a calcitonin-type peptide (ArCT) has been examined in detail in both larval⁶⁴ and adult⁹³ starfish of the species *A. rubens*. Furthermore, pharmacological studies revealed that ArCT acts as muscle relaxant in *A. rubens* and a ortholog of ArCT (PpCT) acts as a muscle relaxant in the starfish species *Patiria pectinifera*⁹³. Calcitonin-related peptides also act as muscle relaxants in vertebrates and therefore it was concluded that this role may be evolutionarily ancient, dating back to the common ancestor of deuterostomes⁹³. In this context, it will be interesting to investigate the actions of AjCT1 and AjCT2 in *A. japonicus* to determine if they also act as myorelaxants.

Pigment-dispersing factor-type precursors (AjPDFP1a and AjPDFP1b). The *A. japonicus* pigment-dispersing factor-type precursor (AjPDFP1a) identified here has been reported previously but was incorrectly annotated as a corticotropin-releasing hormone (CRH)-type precursor³². AjPDFP1a is a 133-residue protein comprising a predicted 21-residue N-terminal signal peptide and two putative PDF-type peptides bounded by dibasic cleavage sites: AjPDF1, a 35-residue peptide, and AjPDF2, a 32-residue peptide with a C-terminal glycine residue that is a potential substrate for amidation (Fig. 2, GenBank accession number MF401990). A second PDF-type neuropeptide precursor in *A. japonicus* (AjPDFP1b) that has not been reported previously was also identified. AjPDFP1b is a 111-residue protein comprising a predicted 21-residue N-terminal signal peptide and two putative PDF-type neuropeptides bounded by dibasic cleavage sites (AjPDF1 and AjPDF2) that are identical to the PDF-type peptides derived from AjPDFP1a (Fig. 2, GenBank accession number MF401991). However, comparison of the sequences of AjPDFP1a and AjPDFP1b reveals that the length of the segment of the precursor between the signal peptide and the first dibasic cleavage site is longer in AjPDFP1a than in AjPDFP1b (Fig. 2). These findings indicate that these two precursors are encoded by transcripts that are alternatively spliced products of

the same gene. The functional significance of the occurrence of two PDF-type precursor *isoforms* in *A. japonicus* is unknown, but it is interesting because it has not been reported in other echinoderms. Furthermore, it is unusual in as much as the neuropeptides derived from the two precursors are the same and it is only what is presumed to be non-bioactive segment of the precursor that differs.

PDF-type neuropeptides were first discovered in crustaceans⁹⁴ but subsequently have been characterised in other arthropods^{95–98} and in other protostomian invertebrates, including nematodes⁹⁹ and lophotrochozoans^{72,100} and an alignment of AjPDF1 and AjPDF2 with PDF-type neuropeptides that have been identified in other taxa is shown in Supplementary Fig. 1K. PDF-type peptides have diverse physiological roles in protostomes, including regulation of pigment migration⁹⁴, circadian patterns of locomotor activity^{96–98}, egg-laying¹⁰¹ and feeding behavior¹⁰². Nothing is known about the physiological roles of PDF-type neuropeptides in echinoderms, so the identification of PDF-type neuropeptides in *A. japonicus* and other echinoderms provides exciting opportunities to address this issue.

Pedal peptide-type precursor 2 (AjPPLNP2). The *A. japonicus* pedal peptide-type precursor (AjPPLNP2) identified here has been reported previously³² and is a 173-residue protein comprising a 21-residue N-terminal signal peptide and five predicted pedal peptide-like neuropeptides: AjPPLN2a (FGNSNMDPLVHSLIGG, 1 copy), AjPPLN2b (FGSSQIMDPLRYSLVSG, 2 copies), AjPPLN2c (FGNSNMDPLMYSMIGG, 1 copy), AjPPLN2d (FGNSNMDPLVHSLISGG, 1 copy), and AjPPLN2e (FGYHPMDPLSNLSMSG, 1 copy), all bounded by dibasic cleavage sites (Fig. 2, GenBank No. MF401980). Three of the peptides (AjPPLN2a, AjPPLN2b, AjPPLN2c) were detected by mass spectrometry in CNR extracts, with the C-terminal glycine residue converted to an amide group in the mature peptides, as expected (Supplementary Fig. 2C–E and Supplementary Fig. 3C–E). Interestingly, pedal peptide-type neuropeptides that have been identified in starfish are not amidated^{42,103}. Likewise the majority of the pedal peptide-type neuropeptides in the sea urchin *S. purpuratus* are not amidated, although two of the peptides derived from SpPPLNP1 are predicted to be amidated and this has been confirmed by mass spectrometry^{31,39}. An alignment of pedal peptide-type neuropeptides from *A. japonicus* with related peptides from other taxa is shown in Supplementary Fig. 1L.

Pedal peptides were originally discovered in molluscs, where they regulate contraction of pedal muscles and cilia beating associated with the foot^{104–106}. Orcokinin-type peptides in arthropods, which are orthologs of molluscan pedal peptides, are involved in regulation of a variety of physiological processes, including circadian and seasonal changes in activity^{107–109}. Recently, it was discovered that pedal peptide-type neuropeptides in starfish act as muscle relaxants^{103,110}. However, nothing is known about the physiological roles of pedal peptide-type neuropeptides in sea cucumbers and therefore this represents an interesting area of enquiry for future work.

Glycoprotein hormone alpha-2-type precursor 1 (AjGPA2P1) and Glycoprotein hormone alpha-2-type precursor 2 (AjGPA2P2). The *A. japonicus* glycoprotein hormone alpha-2-type precursor 1 (AjGPA2P1) has been reported previously³² and is a 132-residue protein comprising a predicted 30-residue N-terminal signal peptide, a potential dibasic cleavage site (residues 34 and 35) followed by a 102-residue polypeptide that contains ten cysteine residues, which is typical for GPA2-type polypeptides (Fig. 3, GenBank accession number MF401993). The *A. japonicus* glycoprotein hormone alpha-2-type precursor 2 (AjGPA2P2) has not been reported previously, but only a partial sequence comprising a predicted 34-residue N-terminal signal peptide, followed by a 55-residue sequence containing four cysteine residues was identified by analysis of CNR transcriptome data (Fig. 3, GenBank accession number MH636350). An alignment of the *A. japonicus* GPA2-type peptides with GPA2-type peptides from other taxa is shown in Supplementary Fig. 1M, with many conserved residues providing evidence of relatedness.

Glycoprotein hormone beta-5-type precursor (AjGPB5P). The *A. japonicus* glycoprotein hormone beta-5-type precursor (AjGPB5P) has been reported previously³² and is a 190-residue protein comprising a predicted 25-residue N-terminal signal peptide followed by a 165-residue sequence; cleavage at a monobasic site (residue 96) would yield a 94-residue polypeptide that shares sequence similarity with other GPB5-type subunits and contains only nine cysteine residues (Fig. 3, GenBank accession number 401994). This is atypical for GPB5-type subunits, which typically have ten cysteine residues but, interestingly, it is consistent with the three GPB5-type hormones identified in another echinoderm species – the starfish *A. rubens*⁴². An alignment of the *A. japonicus* GPB5-type peptide with GPB5-type peptides from other taxa is shown in Supplementary Fig. 1N, with many conserved residues providing evidence of relatedness.

GPA2- and GPB5-type subunits have been identified in both vertebrates and invertebrates¹¹¹. Furthermore, dimerization of the GPA2 and GPB5 subunits forms the hormone thyrostimulin that acts as a ligand for TSH receptors¹¹². GPA2/GPA5-type glycoprotein hormones have also been functionally characterized in insects^{113,114}. Our findings indicate the existence of at least two glycoprotein-type hormones in *A. japonicus*: GPA2a/GPB5 and GPA2b/GPB5. However, currently nothing is known about the physiological roles of GPA2/GPB5-type hormones in echinoderms, so the discovery of GPA2-type and GPA5-type subunits in *A. japonicus* provides an opportunity to address this issue.

Bursicon alpha-type precursors (AjBALPP or AjBAP) and Bursicon beta-type precursor (AjBBLPP or AjBBP). The *A. japonicus* bursicon alpha-type precursor (AjBALPP) and bursicon beta-type precursor (AjBBLPP) have been reported previously³². AjBALPP comprises a predicted 26-residue N-terminal signal peptide, and a bursicon alpha-type neuropeptide containing eleven cysteine residues (Fig. 3, GenBank accession number MH636351). AjBBLPP comprises a predicted 24-residue N-terminal signal peptide, and a bursicon beta-type neuropeptide containing eleven cysteine residues (Fig. 3, GenBank accession number MF401995). Alignments of AjBBLP and AjBALP with bursicon alpha-type and beta-type peptides from other taxa are shown in Supplementary Figs. 1O,P, respectively, with many conserved residues providing evidence of relatedness.

Bursicon was first discovered in insects on account of its effect in causing cuticular tanning¹¹⁵ and subsequent studies on crustaceans have revealed that this neuropeptide is involved in the cuticle hardening and regulation of ecdysis^{116–118}. Precursors of bursicon-type neuropeptides have been identified in at least one species belonging to three echinoderm classes: Echinozoa (e.g. *S. purpuratus*;³¹), Holothurozoa (e.g. *A. japonicus*;³², this study) and Asterozoa (e.g. *A. rubens*;⁴²), but currently nothing is known about the physiological roles of bursicon-type neuropeptides in this phylum.

Insulin-like peptide precursor (AjILPP). An *A. japonicus* insulin-like peptide precursor (AjILPP) was identified here as a 156 amino acid protein comprising sequentially (1) a predicted 35 amino acid N-terminal signal peptide, (2) a 30 amino acid polypeptide comprising three cysteine residues (B-chain), (3) a dibasic cleavage site, (4) a C-peptide domain (residues 68–101), (5) a dibasic cleavage site, and (6) a 53-residue polypeptide comprising five cysteine residues (A-chain) (Fig. 3, GenBank accession number MF401996). The A-chain of AjILP also has the cysteine motif CCxxxCxxxxxxxxC, which is the typical signature for the insulin/insulin-like growth factor (IGF)/relaxin superfamily. However, unlike the relaxin/insulin-like (INSL) subclass, the final residue of the A-chain is not a cysteine. The B-chain of AjILP has a shorter cysteine motif CxxxxxxxxC and lacks the typical relaxin-specific receptor-binding motif¹¹⁹. However, the AjILP A-domain does not extend to a D-domain and E-domain, a feature found in the IGF-type precursor¹²⁰, which suggests that AjILP resembles insulin rather than IGF. Alignments of the AjILP A-chain and B-chain with insulin-like peptide A-chains and B-chains from other taxa are shown in Supplementary Fig. 1Q,R, respectively, with many conserved residues providing evidence of relatedness. These include cysteine residues that are known to form disulphide bridges in insulin-related peptides and it is noteworthy that both the AjILP A-chain and B-chain contain an additional cysteine in comparison with A- and B-chains from other taxa. The presence of these additional cysteine residues in the AjILP A- and B-chains suggests that an additional and atypical disulphide bridge may be formed in the putative mature dimeric peptide derived from AjILPP. Therefore, it will be of interest to determine the structure of the mature bioactive peptide derived from AjILPP.

Peptides of the insulin-relaxin superfamily are implicated in critical physiological processes such as nutrient metabolism, cell proliferation, cell survival, reproduction and aging, and have been identified in a wide range of animal taxa from invertebrates such as nematode worms, mollusks, and insects to vertebrates^{121–124}. In the echinoderms, two members of this superfamily were first identified in the sea urchin *S. purpuratus* (SpILP1 and SpILP2) and it was suggested that SpILP1 was potentially involved in feeding behavior whereas SpILP2 may function as a growth signal during embryogenesis¹²⁵. A relaxin-like gonad-stimulating peptide precursor (ArRGPP), a second relaxin-like peptide precursor (ArRLPP2) and two insulin-like growth factors (ArIGFP1 and ArIGFP2) were also identified in the starfish *A. rubens* recently⁴². Furthermore, consistent with the original discovery of RGP as a gonadotropic peptide in starfish⁴¹, ArRGP triggers spawning and oocyte maturation in *A. rubens*¹²⁶. The identification of AjILP in *A. japonicus* provides a basis for further investigation of the physiological roles of insulin-related peptides in echinoderms.

***A. japonicus* neuropeptide precursors that have thus far only been found in echinoderms.** *L-type SALMFamide precursor (AjL-SALMFaP) and F-type SALMFamide precursor (AjF-SALMFaP).* The *A. japonicus* SALMFamide-type neuropeptide precursors (AjL-SALMFaP and AjF-SALMFaP) have been reported previously^{29,51}. AjL-SALMFaP is a 176 amino acid residue protein comprising three putative neuropeptides (Fig. 4, GenBank accession number MF402000) and AjF-SALMFaP is a 290 residue protein comprising eight putative neuropeptides, two of which have been structurally characterized²⁴ (Fig. 4, GenBank accession number MF401999). Here our mass spectrometric analysis of CNR extracts determined the structure of one of the neuropeptides derived from AjL-SALMFaP – RMGFTGNTGILL-NH₂, (Supplementary Figs. 2F and 3F). Alignments of the peptides derived from AjL-SALMFaP and AjF-SALMFaP with L-type and F-type SALMFamides from other echinoderms are shown in Supplementary Figs. 1S,T, respectively.

Previous studies have investigated the expression and pharmacological activity of neuropeptides derived from the F-type SALMFamide precursor in sea cucumbers, revealing that they act as muscle relaxants^{21,22}. However, the expression pattern and actions of neuropeptides derived from the L-type SALMFamide precursor in sea cucumbers has yet to be examined. Determination of the structure of one of the neuropeptides derived from AjL-SALMFaP, as reported here, provides a basis for further studies on L-type SALMFamides in *A. japonicus*.

AN peptide precursor (AjANPP). A partial sequence of an AN peptide-type neuropeptide precursor in *A. japonicus* (AjANPP) has been reported previously as Ajnp5³². Here analysis of CNR transcriptome data also identified a partial precursor sequence but analysis of genome sequence data enabled determination of the complete amino-acid sequence of AjANPP, which is a 555 residue protein comprising the putative neuropeptide sequences: ANRRFSVG (1 copy), ANRYNALRG (3 copies) and ANRYNALREE (1 copy) (Fig. 4, GenBank accession number MF402001). Interestingly, our mass spectroscopic analysis of CNR extracts revealed the presence of a peptide with the sequence ANRYNALR, which was unexpected because it neither has a C-terminal glycine residue nor a C-terminal amide group (Supplementary Figs. 2G and 3G). One possible explanation for this is that the C-terminal glycine residue has been lost from the mature peptide due the action of carboxypeptidase. Alternatively, this could be an example of atypical neuropeptide precursor processing where cleavage occurs N-terminal to a cleavage site comprising a Gly-Lys dipeptide sequence. Further studies are required to investigate this.

An alignment of the *A. japonicus* AN peptides with AN peptides from other echinoderms is shown in Supplementary Fig. 1U. However, nothing is known about the physiological roles of AN peptides in sea cucumbers or in other echinoderms.

Neuropeptide precursor 9 (Ajnp9). The *A. japonicus* protein Ajnp9 has been reported in a previous study³² and is a 50-residue precursor comprising a 20-residue N-terminal signal peptide followed by a 13-residue sequence (residues 21–33) followed by a glycine (residue 34), which is a potential substrate for C-terminal amidation, and then there is a putative dibasic cleavage site (KR) at residues 35 and 36 (Fig. 4, GenBank accession number MH822300). An alignment of Ajn9 with a related peptide from the sea urchin *S. purpuratus* (Spn9;³) is shown in Supplementary Fig. 1V.

Neuropeptide precursor 11 (Ajnp11). The *A. japonicus* protein Ajnp11 has been reported in a previous study³⁴ and is a 90-residue precursor comprising a 20-residue N-terminal signal peptide followed by a 70-residue sequence (residues 21–90) that contains a putative dibasic cleavage site at residues 33/34 (Fig. 4, GenBank accession number MF402004). The N-terminal region of the protein (residue 21–36) contains five acidic residues (D or E), which indicates that this part of the protein may be an acidic spacer peptide. Accordingly, we propose that it is the 56-residue polypeptide formed by residues 35–90 (Ajnp11) that may be a secreted bioactive neuropeptide. It is noteworthy that the 56-residue polypeptide includes six cysteine residues, which may form up to three intramolecular disulfide bridges. Alternatively, a homodimeric protein could be formed by up to six intermolecular disulfide bridges. An alignment of Ajnp11 with related peptides from other echinoderms is shown in Supplementary Fig. 1W.

Neuropeptide precursor 15 (Ajnp15). The *A. japonicus* protein Ajnp15 has been reported in a previous study³⁴ and is a 102-residue protein comprising a predicted 20-residue N-terminal signal peptide followed by an 82-residue sequence (residues 21–102) that contains a putative dibasic cleavage site (KR) at residues 56 and 57 (Fig. 5, GenBank accession number MF402006). We propose that it is the 45-residue peptide formed by residues 58–102 that may be a secreted bioactive neuropeptide. The presence of six cysteine residues in the 45-residue polypeptide suggests that there may be up to three intramolecular disulfide bridges. Alternatively, a homodimeric protein could be formed with up to six intermolecular disulfide bridges. An alignment of Ajnp15 with related peptides from other echinoderms is shown in Supplementary Fig. 1X.

Neuropeptide precursor 18 (Ajnp18). The *A. japonicus* protein Ajnp18 has been reported previously^{33,34} and is a 121-residue protein comprising a predicted 25-residue N-terminal signal peptide followed by a 96-residue polypeptide sequence (residues 26–121) (Fig. 5, GenBank accession number MF422081). We propose that the 96-residue polypeptide may be a secreted bioactive neuropeptide (Ajnp18). It is noteworthy that Ajnp18 contains eight cysteine residues, which may form up to four intramolecular disulfide bridges. Alternatively, a homodimeric protein could be formed with up to eight intermolecular disulfide bridges. Ajnp18 and related peptides that have been identified in the sea urchin *S. purpuratus* (Spn18;³¹) and in the starfish *A. rubens* (Arn18;⁴²) do not share any apparent sequence similarity with neuropeptides or peptide hormones that have been identified in other phyla. An alignment of Ajnp18 with related peptides from other echinoderms is shown in Supplementary Fig. 1Y.

Neuropeptide precursor 23 (Ajnp23). Ajnp23 has been reported in *A. japonicus* in a previous study³⁴. Ajnp23 is a 128-residue protein comprising a predicted 24-residue N-terminal signal peptide followed by a 104-residue polypeptide sequence (residues 25–128) that contains putative dibasic cleavage sites at residues 82/83, 97/98 and 119/120 (Fig. 5, GenBank accession number MF422083). We propose that it is the 20-residue polypeptide formed by residues 99–118 (Ajnp23) that may be a secreted bioactive neuropeptide. It is noteworthy that Ajnp23 contains two cysteine residues that are separated by four amino acid residues and that may form an intramolecular disulfide bridge. Alternatively, a heterodimeric protein could be formed by up to two intermolecular disulfide bridges. Interestingly, our mass spectrometry analysis confirmed the presence of a peptide (APLADDTAHQVDE) that is bounded by putative dibasic cleavage sites in Ajnp3 and which is located N-terminal to Ajnp23 in the precursor sequence (Supplementary Figs. 2H and 3H). However, as this peptide contains four acidic residues, we speculate that it may be an acidic spacer rather than a secreted bioactive neuropeptide. An alignment of the putative peptides derived from Ajnp23 with related peptides from the starfish *A. rubens*⁴² is shown in Supplementary Fig. 1Z.

Neuropeptide precursor 25 (Ajnp25). The *A. japonicus* protein Ajnp25 was identified here for the first time and named according to its similarity to Arnp25, a putative neuropeptide precursor in the starfish *A. rubens*⁴². Ajnp25 is a 147-residue protein comprising a predicted 30-residue N-terminal signal peptide followed by a 117-residue polypeptide sequence (residues 31–147) that contains a putative dibasic cleavage site at residues 48/49 (Fig. 5, GenBank accession number MF422084). We propose that it is the 17-residue peptide (Ajnp25) formed by residues 31–47 that may be a secreted bioactive neuropeptide. The presence of a C-terminal glycine residue on the peptide suggests that this may be a substrate for amidation. It is also noteworthy that, like Arn25, Ajnp25 contains two cysteine residues, which may form an intramolecular disulfide bridge. Alternatively, a homodimeric protein could be formed by up to two intermolecular disulfide bridges. An alignment of Ajnp25 with Arn25 is shown in Supplementary Fig. 1AA. Ajnp25 represents an interesting candidate neuropeptide for further investigation.

Proteins that are precursors of bioactive peptides that have been discovered previously in *A. japonicus*. **GLRFA precursor (AjGLRFA-P).** The *A. japonicus* protein AjGLRFA-P was identified previously²⁹ and is a 62-residue protein comprising a 22-residue signal peptide and the 5-residue peptide GLRFA (Fig. 6). However, a transcript encoding AjGLRFA-P was not identified in our CNR transcriptome data. GLRFA was originally identified as a component of extracts of *A. japonicus* that causes potentiation of electrically-evoked contractions of the radial longitudinal muscle and contraction of intestinal preparations in this species^{23,24}. Recently, expression analysis of a homolog of AjGLRFA-P in the sea cucumber *Holothuria scabra* revealed that it is expressed in the radial nerve cords and the CNR in this species³⁴. Interestingly, the *H. scabra* GLRFA-type

precursor comprises a peptide sequence (GLLGL) that only shares an N-terminal GL motif with the *A. japonicus* peptide. This contrasts with other regions of these precursors that share higher levels of interspecies sequence similarity (e.g. there are two conserved copies of the tripeptide sequence HPC). Therefore, the possibility remains that the GLRFA peptide in *A. japonicus* and the GLLGL peptide in *H. scabra* may be fragments of larger bioactive peptides.

GN19 precursor (AjGN19P). The *A. japonicus* GN19 precursor has been identified previously (292) and is an 82-residue protein (Fig. 6; GenBank accession number MF422085) comprising a 24-residue signal peptide and the C-terminally located 19-residue GN19 peptide, which was originally identified as a peptide that causes contraction or relaxation of intestinal preparations from *A. japonicus*²⁴. Here our mass spectrometric analysis confirmed that GN19 is present in CNR extracts, providing evidence that it acts as a neuropeptide. Furthermore, mass spectrometric analysis of CNR extracts revealed the presence of a second peptide derived from AjGN19P, GQMMRDMDLLEESL, which is bounded by putative dibasic cleavage sites in the precursor protein (Supplementary Figs. 2I–J and 3I–J). However, it remains to be determined whether this peptide is a bioactive neuropeptide or if it simply functions as an acidic spacer peptide in the precursor. Recently, expression analysis of a homolog of AjGN19P in the sea cucumber *H. scabra* revealed that it is expressed in the radial nerve cords, radial longitudinal muscle and intestine in this species³⁴. However, more detailed analysis of anatomical patterns of expression of GN19-type peptides in sea cucumbers has yet to be performed and therefore this represents an important avenue for future work on this neuropeptide system.

Stichopin precursor (AjStichopinP). The *A. japonicus* stichopin precursor (AjStichopinP) has been identified previously²⁹ and is a 39-residue protein comprising a 22-residue signal peptide and the 17-residue bioactive peptide (Fig. 6, GenBank accession number MH636354). Furthermore, as discussed in the introduction, the expression pattern and bioactivity of stichopin have been examined in detail^{23,24,27,28}. Thus far, stichopin has only been identified in *A. japonicus* and related peptides have as yet not been identified in the other sea cucumber species or other echinoderms. As transcriptome/genome sequence data become available for a variety to sea cucumber species it may be possible to determine the phylogenetic distribution and evolutionary origin of stichopin.

SWYG precursors (Aj_SWYGP-2 and Aj_SWYGP-3). Two SWYG-type neuropeptide precursors in *A. japonicus* were identified in our CNR transcriptome (AjSWYGP-2 and AjSWYGP-3). AjSWYGP-2 is a 158-residue protein comprising a predicted 21-residue N-terminal signal peptide and the putative neuropeptide SWYGTLG bounded by putative monobasic cleavage sites (Fig. 6, GenBank accession number MF422086). Aj_SWYGP-3 is a 152-residue protein comprising a predicted 21-residue N-terminal signal peptide and, bounded by putative monobasic cleavage sites, the peptide SWYGSLASALLG, which was detected in CNR extracts by mass spectrometry with conversion of the C-terminal glycine to an amide (Fig. 6, GenBank accession number MF422087, Supplementary Figs. 2K and 3K).

The peptides SWYG-1 (SWYGSLG), SWYG-2 (SWYGTLG) and SWYG-3 (SWYGSLA) were originally isolated from *A. japonicus* as myoactive peptides^{23,24} and a partial sequence of the SWYG-3 precursor was reported previously²⁹. Here we have determined the sequence of a transcript encoding the full-length protein sequence of the SWYG-3 precursor and have demonstrated that a C-terminally extended and amidated isoform of SWYG-3 (SWYGSLASALLamide) is derived from this precursor. In addition, identification of the gene encoding Aj_SWYGP3 (see Table 2 and Supplementary File 1) demonstrates that this precursor is encoded within the *A. japonicus* genome and it is not a bacterial contaminant, as was speculated previously²⁹. The SWYG-2 precursor presented here has not been reported previously. With the discovery of the SWYG precursor transcripts, there now exist opportunities to investigate the expression of these precursors in *A. japonicus* to gain further insights into the physiological roles of SWYG-type peptides in sea cucumbers.

APHAIRPPSG precursor. The peptide APHAIRPPSG was originally identified as a component of *A. japonicus* body wall that causes inhibition of electrically evoked contractions of the radial longitudinal muscle in this species^{23,24,29}. Here a transcript encoding the precursor of this peptide has been identified for the first time, revealing that it is 56-residue protein comprising a predicted 21-residue N-terminal signal peptide followed by the ten-residue APHAIRPPSG sequence, which is bounded C-terminally by a putative monobasic cleavage site (R) (Fig. 6, GenBank accession number MH636355). Importantly, mass spectrometric analysis revealed the presence of the APHAIRPPSG peptide in CNR extracts, confirming that the C-terminal glycine is not a substrate for amidation and providing additional evidence that this peptide is a neuropeptide in *A. japonicus* (Supplementary Figs. 2L and 3L).

GYWKDLNRYVKAHKT precursor. The peptide GYWKDLNRYVKAHKT was originally identified as a component of *A. japonicus* body wall that causes inhibition of electrically evoked contractions of the radial longitudinal muscle in this species^{23,24,29}. Here a transcript encoding the precursor of this peptide has been identified for the first time, revealing that it is a 161-residue protein comprising a predicted 20-residue N-terminal signal peptide and the fifteen-residue peptide GYWKDLNRYVKAHKT bounded by putative monobasic cleavage sites (Fig. 6, GenBank accession number MF422088). The presence of this transcript in the CNR provides important evidence that the peptide GYWKDLNRYVKAHKT is a neuropeptide in *A. japonicus*.

MPMNPADYFSRGTVYIPTRDS precursor. The peptide MPMNPADYFSRGTVYIPTRDS was originally identified as a component of *A. japonicus* body wall that causes potentiation of electrically evoked contractions of the radial longitudinal muscle in this species^{23,24,29}. Here a transcript encoding the precursor of this peptide has been identified for the first time, revealing that it is a 91-residue protein comprising a predicted

Precursor	length (aa)	CNR transcript GenBank ID	Other transcript GenBank ID	Number of exons	Zhang <i>et al.</i> ⁵ predicted protein GenBank ID	Li <i>et al.</i> ⁶ linkage group	References
<i>Ajholotocin</i>	163	MF401997	HADD01035242	3 (4–5)	—	—	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>AjNGIWIYaP</i>	238	MF401992, AB758561	GAVS01019571	3 (5)	PIK59368	LG9	<i>Kato et al.</i> ²⁶ <i>Reich et al.</i> ⁵⁰ <i>Elphick</i> ³¹ <i>Iwakoshi et al.</i> ²³
<i>AjGnRHP</i>	129	MF401983	—	4	PIK40744	—	
<i>AjCRZP</i>	112	MF401982	HADF01080452	1 (2)	—	—	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>AjCCKP1</i>	163	MF422090, MH636358	—	1 (3)	—	—	
<i>AjCCKP2</i>	175	MH351773	—	—	—	—	
<i>AjOXPI</i>	153	MF401988	HADD01071389	1 (2)	—	—	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>AjOXPI2</i>	138	MF401989	HADD01064569	3 (4)	—	LG1	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>AjLQP</i>	115	MF401981	—	3	—	LG13	<i>Rowe et al.</i> ³²
<i>AjKPP</i>	180	MF401998	HADE01073736	2 (3–4)	—	LG19	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>AjSSP1 or Ajnp19</i>	129	MF401987	HADE01043988	3	—	LG3	<i>Io et al.</i> ⁴⁹ <i>Zandawala et al.</i> ³³ <i>Delroisse et al.</i> ³²
<i>AjSSP2 or Ajnp16</i>	139	MF402007	HADD01032762	2	—	LG17	<i>Io et al.</i> ⁴⁹ <i>Rowe et al.</i> ³²
<i>AjCTP1</i>	147	MF401985	HADE01007184	4 (6)	PIK54200	LG14	<i>Io et al.</i> ⁴⁹ <i>Rowe et al.</i> ³²
<i>AjCTP2</i>	97	MF401986	GAVS01031871	3 (5–6)	PIK54200	LG14	<i>Reich et al.</i> ⁵⁰ <i>Rowe et al.</i> ³²
<i>AjPDFP1a</i>	133	MF401990	—	4	PIK32875	—	<i>Rowe et al.</i> ³²
<i>AjPDFP1b</i>	111	MF401991	—	3	PIK32875	—	
<i>AjPPLNP (AjPPLNP2)</i>	175	MF401980	—	2	PIK41353	LG1, LG12	
<i>AjGPA2P1</i>	132	MF401993	—	3 (5)	PIK58117	LG9, LG10	<i>Rowe et al.</i> ³²
<i>AjGPA2P2</i>	89	MH636350	—	2	PIK58115	LG9	
<i>AjGPB5P</i>	190	MF401994	—	2	—	LG18	
<i>AjBALPP or AjBAP</i>	165	MH636351	—	1 (2)	PIK53087	—	<i>Rowe et al.</i> ³²
<i>AjBBLPP or AjBBP</i>	125	MF401995	—	2 (3)	—	LG1	<i>Rowe et al.</i> ³²
<i>AjILPP - CNR</i>	156	MF401996	—	2 (4)	—	LG18	
<i>AjL-SALMFaP</i>	178	MF402000	—	3	PIK62714	LG7	<i>Elphick et al.</i> ⁵¹
<i>AjF-SALMFaP</i>	290	MF401999	—	3 (5)	PIK42498	LG1	<i>Elphick</i> ²⁹ <i>Kato et al.</i> ²⁶
<i>AjANPP</i>	555	MF402001	—	1 (3–4)	PIK58162	LG20	<i>Rowe et al.</i> ³²
<i>Ajnp9</i>	50		—	1	—	—	<i>Rowe et al.</i> ³²
<i>Ajnp11</i>	90	MF402004	HADD01047817	4	—	—	<i>Suwansa-ard et al.</i> ³⁴
<i>Ajnp15</i>	102	MF402006	HADD01060387	2 (3)	—	—	<i>Suwansa-ard et al.</i> ³⁴
<i>Ajnp18</i>	121	MF422081	HADD01043681	5 (6)	PIK37087	LG20	<i>Suwansa-ard et al.</i> ³⁴ <i>Zandawala et al.</i> ³³
<i>Ajnp23</i>	129	MF422083	HADF01012730	2 (3)	PIK33567, PIK45777	LG22	<i>Io et al.</i> ⁴⁹ <i>Suwansa-ard et al.</i> ³⁴
<i>Ajnp25</i>	147	MF422084	—	3	—	LG3	
<i>AjGLRFAP</i>	62		HADE01083363	2	—	—	<i>Io et al.</i> ⁴⁹ <i>Elphick</i> ²⁹ <i>Iwakoshi et al.</i> ²³ <i>Kato et al.</i> ²⁶
<i>AjGN19P</i>	82	MF422085	—	2 (4)	—	LG5	<i>Elphick</i> ²⁹ <i>Iwakoshi et al.</i> ²³ <i>Kato et al.</i> ²⁶
<i>AjStichopinP</i>	39	MH636354	—	2 (3)	—	—	<i>Elphick</i> ²⁹ <i>Iwakoshi et al.</i> ²³ <i>Kato et al.</i> ²⁶
<i>AjSWYGP-2</i>	159	MF422086	—	2–3	—	LG12	<i>Elphick</i> ²⁹
<i>AjSWYGP-3</i>	153	MF422087	—	2–3	—	LG12	<i>Elphick</i> ²⁹
<i>AjAPHAIRPPSGP</i>	56	MH636355	—	4 (5)	—	—	<i>Elphick</i> ²⁹
<i>AjGYWKDLNDNYVKAHKTP</i>	161	MF422088	—	2	PIK57833	—	<i>Elphick</i> ²⁹
<i>AjMPMNPADYFSRGTVYIPTRDSP</i>	91	MF422089	—	2 (4)	—	LG5	<i>Elphick</i> ²⁹
<i>AjGLRFALP</i>	62	MH636356	—	1 (2)	—	LG5	

Continued

Precursor	length (aa)	CNR transcript GenBank ID	Other transcript GenBank ID	Number of exons	Zhang <i>et al.</i> ⁵ predicted protein GenBank ID	Li <i>et al.</i> ⁶ linkage group	References
<i>Ajnp11LP</i>	120	MF402003	—	2 (3)	—	—	
<i>Ajnp14</i>	51	MH636359	—	1	PIK35546, PIK47457, PIK39346	LG8, LG20	

Table 2. Neuropeptide precursor genes in *Apostichopus japonicus*. The table includes (i) the length of the precursor protein, (ii) accession number for the transcript sequence, (iii) the gene structure with reference to the number of exons (protein coding exons without parentheses; protein-coding exons + non-protein exons in parentheses), (iv) GenBank ID for the protein as predicted from analysis of the genome sequence data reported by Zhang *et al.*, 2017 and (v) location of the gene with respect to linkage groups, as determined by analysis of the genome sequence data reported by Li *et al.* (2018). Citations for papers that have previously reported partial or complete sequences of the *A. japonicus* neuropeptide precursors are also listed.

25-residue N-terminal signal peptide followed by the 21-residue peptide MPMNPADYFSRGTVYIPTRDS, which is bounded C-terminally by a monobasic cleavage site (Fig. 6, GenBank accession number MF422089). Mass spectroscopic analysis of CNR extracts confirmed the presence and structure of the 21-residue peptide, but with oxidation of the N-terminal methionine residues (Supplementary Figs. 2M and 3M). The presence of the precursor transcript and the 21-residue peptide in the CNR provides important evidence that the peptide MPMNPADYFSRGTVYIPTRDS is a neuropeptide in *A. japonicus*.

Other novel putative neuropeptide precursors identified in *A. japonicus*. *A. japonicus* GLRFA precursor-like precursor (*AjGLRFALP*). The *A. japonicus* GLRFA precursor-like precursor (*AjGLRFALP*) has not been reported previously and was identified by BLAST analysis of the CNR transcriptome sequence data using the GLRFA precursor sequence (Fig. 7, GenBank accession number MH636356) as a query. The *AjGLRFALP* precursor has been discussed previously in section 3.3.1. *AjGLRFALP* shares 51% amino acid sequence identity with the GLRFA precursor and is a 62-residue protein comprising a predicted 22-residue N-terminal signal peptide followed by a 40-residue sequence (residues 22–62) that contains a tribasic cleavage site (KKR) at residue 41/42/43 (Fig. 7). The putative neuropeptide derived from *AjGLRFALP* has the sequence RHPCPKRNVHRDHPCPD, with two cysteine residues indicative of a disulphide bridge.

A. japonicus neuropeptide precursor 11-like precursor (*Ajnp11LP*). *Ajnp11LP* is a novel candidate neuropeptide precursor predicted based on analysis of the CNR transcriptome sequence data using the neuropeptide precursor prediction tool NPsearch (<https://rubygems.org/gems/NpSearch>; Moghul *et al.*, in preparation). *Ajnp11LP* is a 120-residue protein comprising a predicted 19-residue N-terminal signal peptide followed by a 101-residue sequence (residues 20–120) that contains a 52-residue putative neuropeptide sequence (residues 69–120). The putative neuropeptide (*Ajn11L*) contains six cysteine residues, which may form up to three intramolecular disulfide bridges (Fig. 7, GenBank accession number MF402003).

A. japonicus neuropeptide precursor 14-like precursor (*Ajnp14LP*). *Ajnp14LP* is a novel candidate neuropeptide precursor predicted based on analysis the CNR transcriptome sequence data using the neuropeptide precursor prediction tool NPsearch. *Ajnp14LP* is a 51-residue protein comprising a predicted 20-residue N-terminal signal peptide followed by a 31-residue putative neuropeptide sequence (residues 21–51). The putative neuropeptide (*Ajn14L*) contains six cysteine residues, which may form up to three intramolecular disulfide bridges. Alternatively, a homodimeric protein could be formed by up to six intermolecular disulfide bridges (Fig. 7, GenBank accession number MH636359).

The structure and genomic location of genes encoding neuropeptide precursors in A. japonicus. The recently reported sequences of the genome of *A. japonicus*^{5,6} have provided a unique opportunity to investigate the sequences and structure of genes encoding neuropeptide precursors that have been identified in this species both here and in previous studies by analysis of transcriptomic sequence data. Thus, comparison of transcript sequences with genomic sequence data has enabled assessment of the accuracy of precursor sequences based on assembled transcriptome sequence data as well as identification of the positions of introns between protein-coding exons and non-coding exons. Furthermore, genomic scaffolds have been mapped to twenty-two linkage groups in *A. japonicus*^{5,6}, consistent with the haploid chromosome number in this species^{77,127}. By analysing data from Li *et al.*⁶ we have determined the location of genes encoding neuropeptide precursors with respect to the twenty-two linkage groups. Table 2 summarises our findings from analysis of *A. japonicus* genomic sequence data and the data that underpins the information presented in Table 2 is included as a Supplementary File.

The neuropeptide precursors identified in *A. japonicus* range in length from just 39 residues (stichopin precursor, which comprises only a signal peptide and stichopin) to 555 residues (AN peptide precursor). However, the majority of precursors comprise 70–190 residues. Our analysis of genomic sequence data enabled identification of genes encoding 43 of the *A. japonicus* neuropeptide precursors and the only neuropeptide precursor for which the gene was not identified was the *A. japonicus* CCK-type precursor2 (*AjCCKP2*).

Analysis of the structure of genes encoding neuropeptide precursors in *A. japonicus* revealed that they comprise between one and five protein-coding exons. Furthermore, the positions of introns with respect to the protein sequence is illustrated in Figs. 1–7 by underlining the amino acid or pairs of amino acids whose codon or codons

are interrupted by an intron in the gene. Determination of the number of 5' and 3' non-coding exons is of course dependent on the quality (i.e. length) of transcript sequences but for many of the neuropeptide precursors it was possible to determine the locations of exons encoding the 5' and/or 3' untranslated regions (UTRs) of transcripts (Table 2; Supplementary File).

Of the 44 proteins identified here as neuropeptide precursors based on our analysis of transcriptome sequence data, only 20 of these were annotated as protein-coding genes in the genome analysis reported by Zhang *et al.*⁵. This probably reflects the limitations of gene prediction methods in identifying genes encoding neuropeptide precursors, which are typically quite short proteins (see above) and which typically exhibit low levels of sequence conservation at the inter-phylum level. Furthermore, some of the predicted proteins from genome annotation were not consistent with our findings; for example, a precursor annotated as a GnRH-like tetrapeptide (PIK4914.1) is a TRH-type precursor. Thus, we have identified 26 neuropeptide precursor genes in *A. japonicus* that had not been previously identified as protein-coding genes. Furthermore, we have corrected the predicted protein sequence for 19 genes that had been identified previously as genes encoding larger or smaller predicted proteins. In some cases there were inconsistencies between transcriptome sequence data and genome sequence data at the level of individual exons. This probably reflects errors in assembly of the repetitive sequence data, which is a common feature of neuropeptide precursors. Two examples worthy of highlighting here are the F-type SALMFamide precursor and the pedal peptide-type precursor.

The sequence of the *A. japonicus* F-type SALMFamide precursor has been reported previously based on analysis of transcriptome sequence data^{40,51} and the sequence determined here from analysis CNR transcriptome data is consistent with that reported previously. However, analysis of genomic sequence data indicates that the F-type SALMFamide precursor may be longer than previously reported. The discrepancy appears to arise from the occurrence in the genome of three tandem copies of the sequence MFGKREDLNGLDKR, which occurs as only a single copy in the precursor sequence based on assembled transcriptome data. This may be due to an error in assembly of transcriptome/genome sequence data and therefore it will be important to determine the sequence of *A. japonicus* F-type SALMFamide precursor by cDNA cloning and sequencing.

The sequence of an *A. japonicus* pedal peptide-type precursor (AjPPLNP2) has been reported previously based on analysis of transcriptome sequence data³². Here analysis CNR transcriptome sequence data revealed a transcript encoding a partial protein sequence and with several differences in the C-terminal region by comparison with the sequence reported by Rowe *et al.*³². Furthermore, analysis of genomic sequence data revealed discrepancies between the transcript sequences and the exon sequences of the gene. As with the F-type SALMFamide precursor, these discrepancies appear to be due to assembly errors arising from the presence of repetitive sequences and therefore it will likewise be important to determine the sequence of *A. japonicus* pedal peptide precursor by cDNA cloning and sequencing.

The availability of data where genomic scaffolds have been mapped to linkage groups in *A. japonicus*⁶ has enabled us to analyse the location of neuropeptide precursor genes at the linkage group (i.e. chromosome) level for the first time in an echinoderm. Based on available scaffold-to-linkage group information only 30 out of 44 precursors were mapped to linkage groups. Of the precursors identified on linkage groups, the genes encoding AjPPLNP2, AjGPA2P and AjTRHP were identified on two different linkage groups and therefore further work will be required to resolve these discrepancies. It is noteworthy that neuropeptide precursor genes are widely distributed across the genome, with at least one neuropeptide precursor gene located on 18 of the 22 linkage groups in *A. japonicus*. Ten of the linkage groups contain 2–5 genes but in most cases these genes are located on separate scaffolds. Of specific interest are instances where two or more neuropeptide precursor genes are located on the same genomic scaffold because this may be indicative of an evolutionary and/or regulatory association. However, only a few instances of this were identified, including the presence of (i) the AjOXP2 and AjNGIWamideP genes on scaffold 84 of the Zhang *et al.*⁵ genome sequence (ii) the AjGPA2P1 and AjGPA2P2 genes on scaffold 122⁵ and scaffold 17⁶ (linkage group 9) and (iii) the AjGLRFAP and AjGLRFALP genes on scaffold 7428 of the Li *et al.*⁶ genome sequence.

Analysis of genomic sequence data also enabled identification of neuropeptide precursor genes that are subject to alternative splicing of transcripts. Thus, consistent with the identification of transcripts encoding two calcitonin-type neuropeptide precursors that share partial sequence identity (see section 3.1.10), analysis of genome sequence data revealed how alternative splicing of four protein-coding exons (exons 1–4) gives rise to two transcript types. Thus, the AjCTP1 transcript comprises protein-coding exons 1, 2, 3 and 4, whereas the AjCTP2 transcript comprises protein-coding exons 1, 2 and 4. Similarly, consistent with the identification of transcripts encoding two PDF-type neuropeptide precursors that share partial sequence identity (see section 3.1.11), analysis of genome sequence data revealed how alternative splicing of four protein-coding exons (exons 1–4) gives rise to two transcript types. Thus, the AjPDFP1a transcript comprises protein-coding exons 1, 2, 3 and 4, whereas the AjPDFP1b transcript comprises protein-coding exons 1, 3 and 4.

Conclusions

We have presented here a comprehensive analysis of neuropeptide precursors in the sea cucumber *A. japonicus* by analysing our novel CNR transcriptome sequence data, proteomic data and recently reported genome sequence data^{5,6}. This has enabled confirmation of the sequences of precursors that have been reported previously as well as identification of novel neuropeptide precursor sequences that have not been identified previously. Analysis of genome sequence data with reference to linkage groups has revealed that neuropeptide precursor genes are widely distributed across the *A. japonicus* genome, providing the first genome-wide analysis of the locations of neuropeptide precursor genes in an echinoderm. The data presented in this paper provide a basis for further studies where i) the sequences of neuropeptide precursor transcripts are confirmed by cDNA cloning and sequencing, ii) the anatomical expression patterns of neuropeptide precursor transcripts and the neuropeptides encoded by them are analysed in *A. japonicus* using mRNA *in situ* hybridisation methods and immunohistochemistry, and iii)

the *in vitro* and *in vivo* pharmacological actions of neuropeptides in *A. japonicus* are examined. Comprehensive application of these experimental approaches to analysis of neuropeptide signalling in *A. japonicus* will complement the growing body of data that has been obtained from experimental analysis of neuropeptide expression and action in the starfish *A. rubens*^{77,87,93,110,128,129} and in other species^{41,103}. Comparison of the physiological roles of neuropeptides in sea cucumbers, starfish and other echinoderms will provide novel insights into the evolution of neuropeptide function in echinoderms. Furthermore, as highlighted in the introduction, determination of neuropeptide function in *A. japonicus* and in other sea cucumbers may provide a basis for novel applications in sea cucumber aquaculture.

References

- Schwerdtner, M. K. & Ferse, S. C. The history of makassan trepang fishing and trade. *PLoS ONE*. **5**, e11346 (2010).
- Li, C. L. *et al.* Status analysis and sustainable development strategy of sea cucumber *Apostichopus japonicus* Selenka aquaculture industry in Shandong Province. *Prog Fish Sci*. **31**, 126–133 (2010).
- Lovatelli, A. Advances in sea cucumber aquaculture and management. FAO Fisheries Technical Paper, Food and Agriculture Organization of the United Nations, Rome, **425** (2004).
- Yang, H., Hamel, J. F. & Mercier, A. The sea cucumber *Apostichopus japonicus*: history, biology and aquaculture (ed. Yang, H.) **xix** (Academic Press, 2015).
- Zhang, X. J. *et al.* The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *Plos Bio*. **15**, e2003790 (2017).
- Li, Y. L. *et al.* Sea cucumber genome provides insights into saponin biosynthesis and aestivation regulation. *Cell Disc*. **4**, 29 (2018).
- Avery, L. *Caenorhabditis elegans* behavioral genetics: where are the knobs? *BMC Biol*. **8**, 69 (2010).
- Holden-Dye, L. & Walker, R. J. The roles of neuropeptides in *Caenorhabditis elegans* including their importance in the regulation of feeding and metabolism. *Protein Pept Lett*. **20**, 636–646 (2013).
- Taghert, P. H. & Nitabach, M. N. Peptide neuromodulation in invertebrate model systems. *Neuron*. **76**, 82–97 (2012).
- Van den Pol, A. N. Neuropeptide transmission in brain circuits. *Neuron*. **4**, 98–115 (2012).
- Jékely, G. *et al.* The long and the short of it - a perspective on peptidergic regulation of circuits and behaviour. *J Exp Biol*. **221**, jeb166710 (2018).
- Douglass, J., Civelli, O. & Herbert, E. Polypeptide gene expression: generation of diversity of neuroendocrine peptides (ed. Douglass, J.) **53**, 665–715 (Annu Rev Biochem, 1984).
- Steiner, D. F. The proprotein convertases. *Curr Opin Chem Biol*. **2**, 31–39 (1998).
- Eipper, B. A., Stoffers, D. A. & Mains, R. E. The biosynthesis of neuropeptides: Peptide alpha-amidation. *Annu Rev Neurosci*. **15**, 57–85 (1992).
- Abraham, G. N. & Podell, D. N. Pyroglutamic acid non-metabolic formation, function in proteins and peptides, and characteristics of the enzymes effecting its removal. *Mol. Cell Biochem*. **38**, 181–190 (1981).
- Mirabeau, O. & Joly, J. S. Molecular evolution of peptidergic signaling systems in bilaterians. *Proc Natl Acad Sci USA*. **110**, e2028–e2037 (2013).
- Jékely, G. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc Natl Acad Sci USA*. **110**, 8702–8707 (2013).
- Elphick, M. R., Mirabeau, O. & Larhammar, D. Evolution of neuropeptide signalling systems. *J Exp Bio*. **221**, jeb151092 (2018).
- Diaz-Miranda, L. *et al.* Characterization of two novel neuropeptides from the sea cucumber *Holothuria glaberrima*. *Biol Bull*. **182**, 241–247 (1992).
- Elphick, M. R., Price, D. A., Lee, T. D. & Thorndyke, M. C. The SALMFamides: a new family of neuropeptides isolated from an echinoderm. *Proc R Soc Lond B*. **243**, 121 (1991).
- Diaz-Miranda, L., Blanco, R. E. & García-Ararrás, J. E. Localization of the heptapeptide GFSKLYFamide in the sea cucumber *Holothuria glaberrima* (Echinodermata): a light and electron microscopic study. *J Comp Neurol*. **352**, 626–640 (1995).
- Diaz-Miranda, L. & García-Ararrás, J. E. Pharmacological action of the heptapeptide GFSKLYFamide in the muscle of the sea cucumber *Holothuria glaberrima* (Echinodermata). *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol*. **110**, 171–176 (1995).
- Iwakoshi, E. *et al.* Comparative aspects of structure and action of bioactive peptides isolated from the sea cucumber *Stichopus japonicus*. In *Peptide chemistry*. 261–264 (1995).
- Ohtani, M., Iwakoshi, E., Muneoka, Y., Minakata, H. & Nomoto, K. Isolation and characterization of bioactive peptides from the sea cucumber, *Stichopus japonicus*. In *Peptide Science-Present and Futur* **2002**, 419–420 (1999).
- Inoue, M. *et al.* Localization of the neuropeptide NGIWYamide in the holothurian nervous system and its effects on muscular contraction. *Proc R Soc Lond B*. **266**, 993–1000 (1999).
- Kato, S. *et al.* Neuronal peptides induce oocyte maturation and gamete spawning of sea cucumber. *Apostichopus japonicus*. *Dev Biol*. **326**, 169–176 (2009).
- Birenheide, R. *et al.* Peptides controlling stiffness of connective tissue in sea cucumbers. *Biol Bull*. **194**, 253–259 (1998).
- Tamori, M. *et al.* Stichopin-containing nerves and secretory cells specific to connective tissues of the sea cucumber. *Proc R Soc Lond B*. **274**, 2279–2285 (2007).
- Elphick, M. R. The protein precursors of peptides that affect the mechanics of connective tissue and/or muscle in the echinoderm *Apostichopus japonicus*. *PLoS ONE*. **7**, e44492 (2012).
- Du, H. *et al.* Transcriptome sequencing and characterization for the sea cucumber *Apostichopus japonicus* (Selenka, 1867). *Plos One*. **7**, e33311 (2012).
- Rowe, M. L. & Elphick, M. R. The neuropeptide transcriptome of a model echinoderm, the sea urchin *Strongylocentrotus purpuratus*. *Gen Comp Endocrinol*. **179**, 331–344 (2012).
- Rowe, M. L., Achhala, S. & Elphick, M. R. Neuropeptides and polypeptide hormones in echinoderms: new insights from analysis of the transcriptome of the sea cucumber *Apostichopus japonicus*. *Gen Comp Endocrinol*. **197**, 43–55 (2014).
- Zandawala, M. *et al.* Discovery of novel representatives of bilaterian neuropeptide families and reconstruction of neuropeptide precursor evolution in ophiuroid echinoderms. *Open Biol*. **7**, 170129 (2017).
- Suwansa-arda, S. *et al.* Transcriptomic discovery and comparative analysis of neuropeptide precursors in sea cucumbers (Holothuroidea). *Peptides*. **99**, 231–240 (2018).
- Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. **29**, 644–652 (2011).
- Priyam, A. *et al.* Sequenceserver: a modern graphical user interface for custom BLAST databases. *Biorxiv*, <https://doi.org/10.1101/033142> (2015).
- Burke, R. D. *et al.* A genomic view of the sea urchin nervous system. *Dev Biol*. **300**, 434–460 (2006).
- Elphick, M. R. & Rowe, M. L. NGFFFamide and echinotocin: structurally unrelated myoactive neuropeptides derived from neurophysin-containing precursors in sea urchins. *J Exp Biol*. **212**, 1067–1077 (2009).
- Menschaert, G. *et al.* A hybrid, de novo based, genome-wide database search approach applied to the sea urchin neuropeptidome. *J Proteome Res*. **9**, 990–996 (2010).

40. Elphick, M. R. & Thorndyke, M. C. Molecular characterisation of SALMFamide neuropeptides in sea urchins. *J Exp Biol.* **208**, 4273–4282 (2005).
41. Mita, M. *et al.* A relaxin-like peptide purified from radial nerves induces oocyte maturation and ovulation in the starfish, *Asterina pectinifera*. *Proc Natl Acad Sci USA* **106**, 9507–9512 (2009).
42. Semmens, D. C. *et al.* Transcriptomic identification of starfish neuropeptide precursors yields new insights into neuropeptide evolution. *Open Biol.* **6**, 150224 (2016).
43. Bendtsen, J. D., Nielsen, H., von, H. G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* **340**, 783–795 (2004).
44. Liu, F. *et al.* *In silico* identification of new secretory peptide genes in *Drosophila melanogaster*. *Mol Cell Proteomics.* **5**, 510–522 (2006).
45. Seidah, N. G. & Chretien, M. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* **848**, 45–62 (1999).
46. Veenstra, J. A. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch Insect Biochem Physiol.* **43**, 49–63 (2000).
47. Moghul, I. *et al.* NpHMMer and NpSearch: tools to identify neuropeptide precursors. In preparation.
48. Baldin, C. *et al.* Comparative proteomics of a Tor inducible *Aspergillus fumigatus* mutant reveals involvement of the Tor kinase in iron regulation. *Proteomics.* **15**, 2230–2243 (2015).
49. Jo, J. *et al.* Comparative transcriptome analysis of three color variants of the sea cucumber *Apostichopus japonicus*. *Mar Genom.* **28**, 21–24 (2016).
50. Reich, A., Dunn, C., Akasaka, K. & Wessel, G. Phylogenomic analyses of Echinodermata support the sister groups of Asterozoa and Echinozoa. *PLoS ONE.* **10**, 1–11 (2015).
51. Elphick, M. R., Achhala, S. & Martyniuk, N. The evolution and diversity of SALMFamide neuropeptides. *PLoS ONE.* **8**, e59076 (2013).
52. Delroisse, J., Malfefet, J. & Flammang, P. De Novo Adult Transcriptomes of Two European Brittle Stars: Spotlight on Opsin-Based Photoreception. *PLoS ONE.* **11**, e0152988 (2016).
53. Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct.* **3**, 20 (2008).
54. Kiryutin B., Souvorov A. & Tatusova T. ProSplign—protein to genomic alignment tool. In Proc. 11th Annual International Conference in Research in Computational Molecular Biology, San Francisco, USA (2007)
55. Sawyer, W. H. Evolution of neurohypophysial hormones and their receptors. *Fed Proc.* **36**, 1842–1847 (1977).
56. Hruby, V. J., al-Obeidi, F. & Kazmierski, W. Emerging approaches in the molecular design of receptors elective peptide ligands: conformational, topographical and dynamic considerations. *Biochem J.* **268**, 249–262 (1990).
57. Urano, A., Hyodo, S. & Suzuki, M. Molecular evolution of neurohypophysial hormone precursors. *Prog Brain Res.* **92**, 39–46 (1992).
58. Hoyle, C. H. Neuropeptide families and their receptors: evolutionary perspectives. *Brain Res.* **848**, 1–25 (1999).
59. Kawada, T., Sekiguchi, T., Itoh, Y., Ogasawara, M. & Satake, H. Characterization of a novel vasopressin/oxytocin superfamily peptide and its receptor from an ascidian, *Ciona intestinalis*. *Peptides.* **29**, 1672–1678 (2008).
60. Ukena, K., Iwakoshi-Ukena, E. & Hikosaka, A. Unique form and osmoregulatory function of a neurohypophysial hormone in a urochordate. *Endocrinology.* **149**, 5254–5261 (2008).
61. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature.* **453**, 1064–1071 (2008).
62. Elphick, M. R. NG peptides: a novel family of neurophysin-associated neuropeptides. *Gene.* **458**, 20–26 (2010).
63. Beets, I., Temmerman, L., Janssen, T. & Schoofs, L. Ancient neuromodulation by vasopressin/oxytocin related peptides. *Worm.* **2**, e24246 (2013).
64. Mayorova, T. D. *et al.* Localization of Neuropeptide Gene Expression in Larvae of an Echinoderm, the Starfish *Asterias rubens*. *Front Neurosci.* **10**, 553 (2016).
65. Semmens, D. C. Discovery of sea urchin NGFFamide receptor unites a bilaterian neuropeptide family. *Open Biol.* **5** (2015).
66. Zandawala, M., Tian, S. & Elphick, M. R. The evolution and nomenclature of GnRH-type and corazonin-type neuropeptide signaling systems. *Gen. Comp. Endocr.* **264**, 64–77 (2018).
67. Tian, S. *et al.* Urbilaterian origin of paralogous GnRH and corazonin neuropeptide signalling pathways. *Sci Rep.* **6** (2016).
68. Kim, D. K. *et al.* Revisiting the evolution of gonadotropin-releasing hormones and their receptors in vertebrates: secrets hidden in genomes. *Gen Comp Endocrinol.* **170**, 68–78 (2011).
69. Iwakoshi, E. *et al.* Isolation and characterization of a GnRH-like peptide from *Octopus vulgaris*. *Biochem Biophys Res Commun.* **291**, 1187–1193 (2002).
70. Gorbman, A. & Sower, S. A. Evolution of the role of GnRH in animal (metazoan) biology. *Gen Comp Endocrinol.* **134**, 207–213 (2003).
71. Tsai, P. S. Gonadotropin-releasing hormone in invertebrates: structure, function, and evolution. *Gen Comp Endocrinol.* **148**, 48–53 (2006).
72. Veenstra, J. A. Neuropeptide evolution: neurohormones and neuropeptides predicted from the genomes of *Capitella teleta* and *Helobdella robusta*. *Gen Comp Endocrinol.* **171**, 160–175 (2011).
73. Veenstra, J. A. Isolation and structure of corazonin, a cardioactive peptide from the American cockroach. *FEBS Lett.* **250**, 231–234 (1989).
74. Kim, Y. J. *et al.* Corazonin receptor signaling in ecdysis initiation. *Proc Natl Acad Sci USA* **101**, 6704–6709 (2004).
75. Verlinden, H., Badisco, L., Marchal, E., Van, W. P. & Vanden, B. J. Endocrinology of reproduction and phase transition in locusts. *Gen Comp Endocrinol.* **162**, 79–92 (2009).
76. Kubrak, O. I., Lushchak, O. V., Zandawala, M. & Nassel, D. R. Systemic corazonin signalling modulates stress responses and metabolism in *Drosophila*. *Open Biol.* **6** (2016).
77. Tian, S., Egertová, M. & Elphick, M. R. Functional Characterization of Paralogous Gonadotropin-Releasing Hormone-Type and Corazonin-Type Neuropeptides in an Echinoderm. *Front Endocrinol.* **8**, 259 (2017).
78. Yu, N. & Smagghe, G. CCK(-like) and receptors: structure and phylogeny in a comparative perspective. *Gen Comp Endocrinol.* **209**, 74–81 (2014).
79. García-Arriarás, J. E., Torres-Avillán, I. & Ortíz-Miranda, S. Cells in the intestinal system of holothurians (Echinodermata) express cholecystokinin-like immunoreactivity. *Gen Comp. Endocrinol.* **83**, 233–242 (1991).
80. Bílek, R., Bičíková, M. & Šafařík, L. TRH-like peptides. *Physiol Res.* **60**, 207–215 (2011).
81. Galas, L. *et al.* TRH acts as a multifunctional hypophysiotropic factor in vertebrates. *Gen Comp Endocrinol.* **164**, 40–50 (2009).
82. Van, S. E. *et al.* Evolutionarily conserved TRH neuropeptide pathway regulates growth in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **114**, E4065–E4074 (2017).
83. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell.* **92**, 573–585 (1998).
84. De, L. L. *et al.* The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc Natl Acad Sci USA* **95**, 322–327 (1998).
85. Sakurai, T. The neural circuit of orexin (hypocretin): maintaining sleep and wakefulness. *Nat Rev Neurosci.* **8**, 171–181 (2007).

86. Matsuda, K., Azuma, M. & Kang, K. S. Orexin system in teleost fish. *Vitam Horm.* **89**, 341–361 (2012).
87. Yañez-Guerra, L. A. *et al.* Discovery and functional characterisation of a luqin-type neuropeptide signalling system in a deuterostome. *Sci Rep.* **8**, 7220 (2018).
88. Oakley, A. E., Clifton, D. K. & Steiner, R. A. Kisspeptin signaling in the brain. *Endocr Rev.* **30**, 713–743 (2009).
89. Tostivint, H., Lihmann, I. & Vaudry, H. New insight into the molecular evolution of the somatostatin family. *Mol Cell Endocrinol.* **286**, 5–17 (2008).
90. Viollet, C. *et al.* Somatostatinergic systems in brain: networks and functions. *Mol Cell Endocrinol.* **286**, 75–87 (2008).
91. Chang, C. L., Roh, J. & Hsu, S. Y. Intermedin, a novel calcitonin family peptide that exists in teleosts as well as in mammals: a comparison with other calcitonin/intermedin family peptides in vertebrates. *Peptides.* **25**, 1633–1642 (2004).
92. Furuya, K. *et al.* Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. *Proc Natl Acad Sci USA* **97**, 6469–6474 (2000).
93. Cai, W. G. *et al.* Biochemical, Anatomical, and Pharmacological Characterization of Calcitonin-Type Neuropeptides in Starfish: Discovery of an Ancient Role as Muscle Relaxants. *Front Neurosci.* **12**, 382 (2018).
94. Fernlund, P. Structure of a light-adapting hormone from the shrimp, *Pandalus borealis*. *Biochim Biophys Acta.* **439**, 17–25 (1976).
95. Rao, K. R. & Riehm, J. P. Pigment-dispersing hormones. *Ann NY Acad Sci.* **680**, 78–88 (1993).
96. Bahn, J. H., Lee, G. & Park, J. H. Comparative analysis of Pdf-mediated circadian behaviors between *Drosophila melanogaster* and *D. virilis*. *Genetics.* **181**, 965–975 (2009).
97. Lee, C. M., Su, M. T. & Lee, H. J. Pigment dispersing factor: an output regulator of the circadian clock in the German cockroach. *J Biol Rhythms.* **24**, 35–43 (2009).
98. Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C. & Taghert, P. H. A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in *Drosophila*. *Cell.* **99**, 791–802 (1999).
99. Janssen, T. *et al.* Discovery and characterization of a conserved pigment dispersing factor-like neuropeptide pathway in *Caenorhabditis elegans*. *J Neurochem.* **111**, 228–241 (2009).
100. Veenstra, J. A. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. *Gen Comp Endocrinol.* **167**, 86–103 (2010).
101. Meelkop, E. *et al.* PDF receptor signaling in *Caenorhabditis elegans* modulates locomotion and egg-laying. *Mol Cell Endocrinol.* **361**, 232–240 (2012).
102. Li, L. *et al.* Cerebrin prohormone processing, distribution and action in *Aplysia californica*. *J Neurochem.* **77**, 1569–1580 (2001).
103. Kim, C. H. *et al.* Identification of a Novel Starfish Neuropeptide That Acts as a Muscle Relaxant. *J Neurochem.* **137**, 33–45 (2016).
104. Lloyd, P. E. & Connolly, C. M. Sequence of pedal peptide: a novel neuropeptide from the central nervous system of *Aplysia*. *J Neurosci.* **9**, 312–317 (1989).
105. Hall, J. D. & Lloyd, P. E. Involvement of pedal peptide in locomotion in *Aplysia*: modulation of foot muscle contractions. *J Neurobiol.* **21**, 858–868 (1990).
106. Longley, R. D. & Peterman, M. Neuronal control of pedal sole cilia in the pond snail *Lymnaea stagnalis* appressa. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol.* **199**, 71–86 (2013).
107. Hofer, S. & Homberg, U. Evidence for a role of orcokinin-related peptides in the circadian clock controlling locomotor activity of the cockroach *Leucophaea maderae*. *J Exp Biol.* **209**, 2794–2803 (2006).
108. Soehler, S., Stengl, M. & Reischig, T. Circadian pacemaker coupling by multi-peptidergic neurons in the cockroach *Leucophaea maderae*. *Cell Tissue Res.* **343**, 559–577 (2011).
109. Wei, H. & Stengl, M. Light affects the branching pattern of peptidergic circadian pacemaker neurons in the brain of the cockroach *Leucophaea maderae*. *J Biol Rhythms.* **26**, 507–517 (2011).
110. Lin, M., Egertová, M., Zamprnio, C. G., Jones, A. M. & Elphick, M. R. Pedal peptide/orcokinin-type neuropeptide signaling in a deuterostome: The anatomy and pharmacology of starfish myorelaxant peptide in *Asterias rubens*. *J Comp Neurol.* **525**, 3890–3917 (2017).
111. Park, J. I., Semyonov, J., Chang, C. L. & Hsu, S. Y. Conservation of the heterodimeric glycoprotein hormone subunit family proteins and the LGR signaling system from nematodes to humans. *Endocrine.* **26**, 267–276 (2005).
112. Nakabayashi, K. *et al.* Thyrostimulin, a heterodimer of two new human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. *J Clin Invest.* **109**, 1445–1452 (2002).
113. Sellami, A., Agricola, H. J. & Veenstra, J. A. Neuroendocrine cells in *Drosophila melanogaster* producing GPA2/GPB5, a hormone with homology to LH, FSH and TSH. *Gen Comp Endocrinol.* **170**, 582–588 (2011).
114. Paluzzi, J. P., Vanderveken, M. & O'Donnell, M. J. The heterodimeric glycoprotein hormone, GPA2/GPB5, regulates ion transport across the hindgut of the adult mosquito, *Aedes aegypti*. *PLoS ONE.* **9**, e86386 (2014).
115. Fraenkel, G., Hsiao, C. & Seligman, M. Properties of bursicon: an insect protein hormone that controls cuticular tanning. *Science.* **151**, 91–93 (1966).
116. Wilcockson, D. C. & Webster, S. G. Identification and developmental expression of mRNAs encoding putative insect cuticle hardening hormone, bursicon in the green shore crab *Carcinus maenas*. *Gen Comp Endocrinol.* **156**, 113–125 (2008).
117. Chung, J. S., Katayama, H. & Dirksen, H. New functions of arthropod bursicon: inducing deposition and thickening of new cuticle and hemocyte granulation in the blue crab, *Callinectes sapidus*. *PLoS ONE.* **7**, e46299 (2012).
118. Webster, S. G., Wilcockson, D. C., Mrinalini & Sharp, J. H. Bursicon and neuropeptide cascades during the ecdysis program of the shore crab, *Carcinus maenas*. *Gen Comp Endocrinol.* **182**, 54–64 (2013).
119. Bullesbach, E. E. & Schwabe, C. The relaxin receptor-binding site geometry suggests a novel gripping mode of interaction. *J Biol Chem.* **275**, 35276–35280 (2000).
120. Reinecke, M. & Collet, C. The phylogeny of the insulin-like growth factors. *Int Rev Cytol.* **183**, 1–94 (1998).
121. Claeys, I. *et al.* Insulin-related peptides and their conserved signal transduction pathway. *Peptides.* **23**, 807–816 (2002).
122. Antonova, Y., Arik, A. J., Moore, W., Riehle, M. R. & Brown, M. R. Insulin-like peptides: structure, signaling, and function. *Insect Endocrinology.* 63–92 (2012).
123. Grönke, S., Clarke, D. F., Broughton, S., Andrews, T. D. & Partridge, L. Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet.* **6**, e1000857 (2010).
124. Garofalo, R. S. Genetic analysis of insulin signaling in *Drosophila*. *Trends Endocrinol Metab.* **13**, 156–162 (2002).
125. Perillo, M. & Arnone, M. I. Characterization of insulin-like peptides (ILPs) in the sea urchin *Strongylocentrotus purpuratus*: Insights on the evolution of the insulin family. *Gen Comp Endocrinol.* **205**, 68–79 (2014).
126. Lin, M. *et al.* Cellular Localization of Relaxin-Like Gonad-Stimulating Peptide Expression in *Asterias Rubens*: New Insights Into Neurohormonal Control of Spawning in Starfish. *J Comp Neurol.* **525**, 1599–1617 (2017).
127. Okumura, S. I. *et al.* Chromosome number and telomere sequence mapping of the Japanese sea cucumber *Apostichopus japonicus*. *Fisheries Sci.* **75**, 249–251 (2009).
128. Semmens, D. C. *et al.* Discovery of a novel neurophysin-associated neuropeptide that triggers cardiac stomach contraction and retraction in starfish. *J Exp Biol.* **216**, 4047–4053 (2013).
129. Tinoco, A. B. *et al.* Characterization of NGFFYamide Signaling in Starfish Reveals Roles in Regulation of Feeding Behavior and Locomotory Systems. *Front Endocrinol.* **9**, 507 (2018).

Acknowledgements

We thank Jingjie PTM Biolab for excellent technical assistance. This research was supported by a grant (No. 41676124) from the National Natural Science Foundation China.

Author Contributions

M.C. and M.R.E. originally conceived the study. M.C., A.T., Y.Z. and M.R.E. contributed to the acquisition of data. M.C., A.T., K.B.S. and M.R.E. contributed to the interpretation of the data. M.C. and M.R.E. wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45271-3>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019