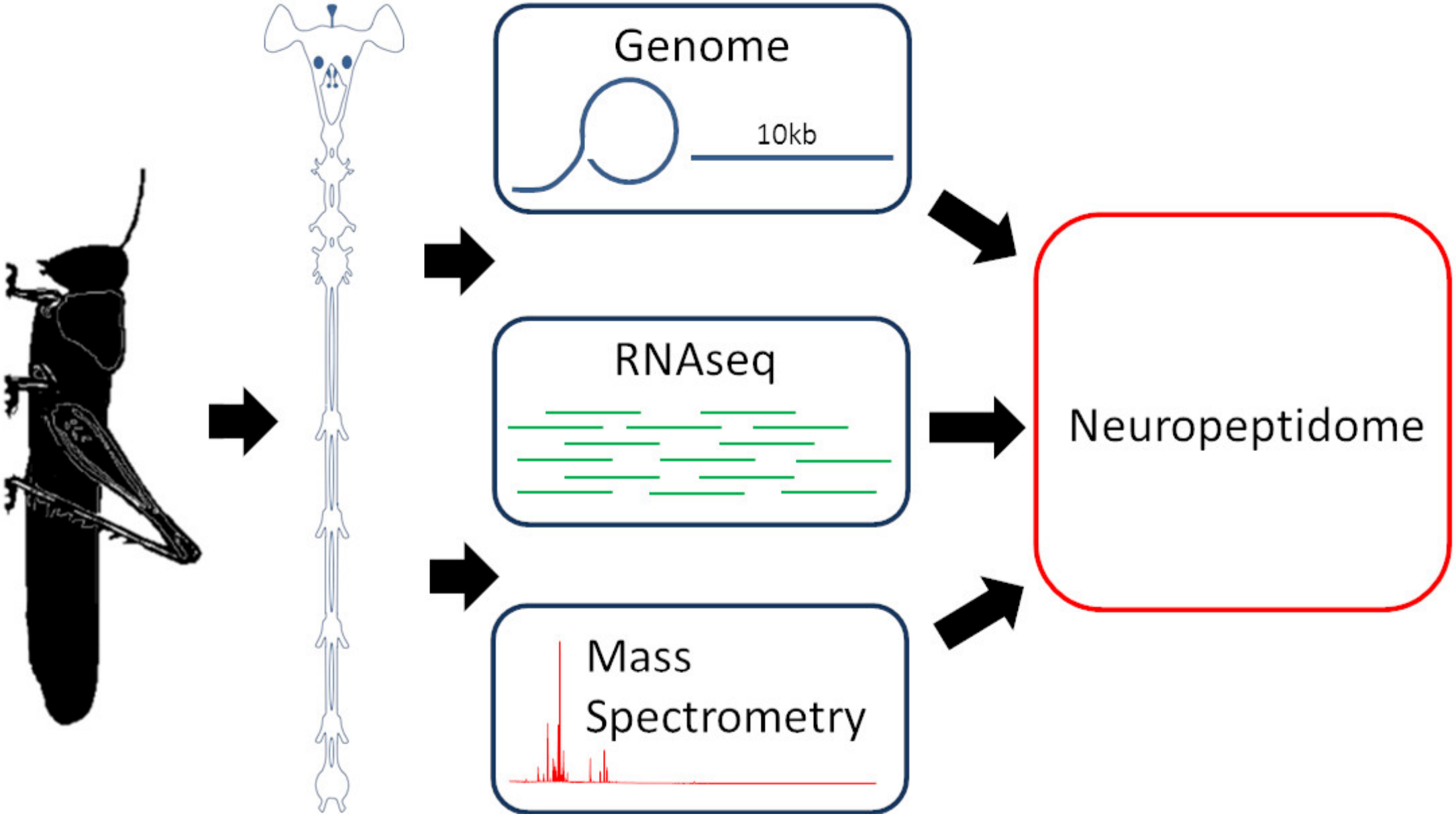
Schistocerca neuropeptides – An update

Peer-reviewed author version

Ragionieri, Lapo; VERDONCK, Rik; Verlinden, Heleen; Marchal, Elisabeth; Vanden Broeck, Jozef & Predel, Reinhard (2022) Schistocerca neuropeptides – An update. In: JOURNAL OF INSECT PHYSIOLOGY, 136 (Art N° 104326).

DOI: 10.1016/j.jinsphys.2021.104326
Handle: http://hdl.handle.net/1942/37403


Genome
10kb
RNAseq
Mass
Spectrometry
Neuropeptidome

Schistocerca neuropeptides - An update

Lapo Ragionieri[a], Rik Verdonck[bc], Heleen Verlinden[b], Elisabeth Marchal[b], Jozef Vanden Broeck[b], Reinhard Predel[a]

[a] University of Cologne, Department of Biology, Institute for Zoology, Zülpicher Str. 47b, 50674, Cologne, Germany

[b] Division of Animal Physiology and Neurobiology, Zoological Institute, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

[c] Centre for Environmental Sciences, Environmental Biology, Hasselt University, Diepenbeek, Belgium

Corresponding authors: lapo.ragionieri@uni-koeln-de, jozef.vandenbroeck@kuleuven.be, rpredel@uni-koeln.de.

Abstract

We compiled a comprehensive list of 67 precursor genes encoding neuropeptides and neuropeptide-like peptides using the *Schistocerca gregaria* genome and several transcriptome datasets. 11 of these 67 precursor genes have alternative transcripts, bringing the total number of *S. gregaria* precursors identified in this study to 81. Based on this precursor information, we used different mass spectrometry approaches to identify the putative mature, bioactive peptides processed in the nervous system of *S. gregaria*. The thereby generated dataset for *S. gregaria* confirms significant conservation of the entire neuropeptidergic gene set typical of insects and also contains precursors typical of Polyneoptera only. This is in striking contrast to the substantial losses of peptidergic systems in some holometabolous species. The neuropeptidome of *S. gregaria*, apart from species-specific sequences within the known range of variation, is quite similar to that of *Locusta migratoria* and even to that of less closely related Polyneoptera. With the *S. gregaria* peptidomics data presented here, we have thus generated a very useful source of information that could also be relevant for the study of other Polyneoptera species.

Keywords: Locust, mass spectrometry, peptidomics, genome, transcriptome

1. INTRODUCTION

Neuropeptides are structurally highly diverse signaling molecules involved in the regulation of most physiological functions in Metazoa. As such, they participate in intercellular information transfer from neurotransmission to intrinsic or extrinsic neuromodulation. For peptides involved in extrinsic neuromodulation, the term peptide hormones is alternatively used (e.g., Nässel and Zandawala, 2020), but the general features of precursor processing, peptide release, and receptor activation do not differ for peptides acting within the nervous system and those acting as hormones. The majority of neuropeptides activate peptide-specific G-protein coupled receptors and quite a few of these peptidergic systems have a very ancient origin that can be traced back to the early evolution of Metazoa (e.g., Elphick et al., 2018, Jékely, 2013, Vanden Broeck, 1996). A special focus in

neuropeptide research has always been on insects (Kopeć, 1922, Scharrer, 1987, Starratt and Brown, 1975) and many neuropeptides were first described from Polyneoptera such as cockroaches (Holman et al., 1991, Predel et al., 2001) and locusts (Schoofs et al., 1993). The physiology of these large insects was well studied at the time and established bioassays were employed to isolate numerous bioactive compounds. Particular attention was paid to locust neuropeptides and their role in regulating physiology, including behavior (Schoofs et al., 2017); driven in part by the need to seek for environmentally sound strategies to limit outbreaks of these pests.

In recent years, the increasing availability of genomic and transcriptomic data has provided comprehensive sequence information on neuropeptide genes in many insects (e.g., non-pterygote Hexapoda, Derst et al., 2016; Polyneoptera, Bläser and Predel, 2020; parasitoid wasps, Chang et al., 2018; Coleoptera, Veenstra, 2019), and the fruit fly *Drosophila melanogaster* has become a preferred model organism for studying neuropeptide function (Nässel and Zandawala, 2019). However, due to their relatively large body size, locusts have remained in the spotlight as classic models for physiological and neurobiological research. This continued interest is further reinforced by the devastating consequences of the largest locust outbreaks since many decades, which were observed in recent years on different continents. While, until recently, genomic analyses were hampered by the huge genome size of locust species, other 'omics' approaches, such as peptidomics (Clynen et al., 2001, Verdonck et al., 2016) and transcriptomics (Badisco et al., 2011, Chen et al., 2010), have been successfully applied, and RNA interference has been shown to be a very robust tool for functional, reverse genetic studies (Santos et al., 2014).

Following the release of the *Locusta migratoria* genome sequence (Wang et al., 2014), Veenstra (2014) compiled a list of predicted precursors for neuropeptides and neuropeptide-like substances. Some of these precursors were incomplete at that time. Later, Hou et al. (2015) revised this list and performed first experiments targeting development-specific, tissue-specific, and phase-specific (gregarious vs. solitarious) expression profiles of neuropeptide precursor genes. Yet, there is little biochemical evidence of the many newly proposed peptides for *L. migratoria*, and the situation is even less favorable for *Schistocerca gregaria*, whose genome has only recently been published (Verlinden et al., 2020). Therefore, our current knowledge regarding the biochemically confirmed neuropeptidome of locusts, i.e., the knowledge of mature peptides resulting from precursor processing, is largely identical to what we knew before the sequencing of locust genomes (Clynen and Schoofs, 2009). However, many more neuropeptides can be expected based on the precursor information obtained from genomes and/or transcriptomes. Other polyneopteran insects such as the American cockroach *Periplaneta americana* (Zeng et al. 2021) and the stick insect *Carausius morosus* (Liessem et al., 2018) have been much better studied in this regard. In the present study, using the *S. gregaria* genome (Verlinden et al., 2020) and several transcriptome datasets, we compiled a comprehensive list of precursors for neuropeptides and neuropeptide-like substances. Based on this information, we used different mass spectrometry approaches to identify the mature and likely bioactive peptides processed in the nervous system of *S. gregaria* and thereby generated an updated list of locust neuropeptides which can be used as reference in future experiments.

2. Materials and Methods

2.1 Insects

Adults and juveniles of *S. gregaria* belonged to a colony originally from Nigeria and maintained since 1985 in the laboratory at KU Leuven, Belgium. Both solitarious and gregarious animals were reared at constant temperature (32 ± 1 °C), with day/night cycle of 13/11 hours, and relative humidity between 40% and 60%. Solitarious animals were reared in visual and olfactory isolation while gregarious animals were reared under crowded conditions. Animals were fed daily with fresh cabbage leaves and dry oat flakes. The animals used for genome sequencing were inbred for seven generations (Verlinden et al., 2020).

2.2 Transcriptome sequencing, *de novo* assembly of transcriptome nucleotide sequences, and quality control

Total RNA was extracted from adults and juveniles of both sexes reared under solitary and gregarious conditions using two protocols. In protocol 1 the CNS of one adult female and 5th instar juvenile reared under gregarious conditions were merged and extracted using TRIzol (Thermo Fisher Scientific, Darmstadt, Germany) following the manufacturer's recommendations (Bio project *pending*). Libraries were prepared starting from 1 µg of total RNA with the Illumina® TruSeq® stranded RNA sample preparation Kit. Libraries were validated and quantified using the Agilent 2100 Bioanalyzer. Sequencing was done using an Illumina TruSeq PE Cluster Kit v3 and an Illumina TruSeq SBS Kit v3 - HS on an Illumina HiSeq 4000 sequencer with a paired-end. In protocol 2 were performed 14 RNA extractions including different tissues (Bio project *pending*): cerebral ganglia, gnathal ganglia, thoracic ganglia, flight muscle, cuticle, antenna, fat body, salivary gland, male and female gonads and eggs (Bio project *pending*). Samples were homogenized in 2 ml MagNA lyser Green Beads vials with 1 ml QIAzol lysis reagent buffer (6 000 x g for 30 seconds). Next, total RNA was extracted using QIAGEN RNeasy Lipid Tissue Mini kit (Company information - Rik) following standard protocols and including the extra DNase digestion step. Samples were pooled in equimolar concentrations to finally yield a master sample with a concentration of 627 ng/µl. Part of the master sample was further enriched in mRNA using the Illustra QuickPrep Micro mRNA Purification Kit (GE Healthcare) 5 libraries were prepared for sequencing: one starting from total RNA using the Illumina Truseq protocol without *in vitro* normalisation, two starting from purified mRNA using the Illumina Truseq protocol (one with and one without *in vitro* normalisation step) and two using the Clontech SMARTer kit (one with and one without *in vitro* normalisation step). Samples were sequenced on two lanes of Illumina HiSeq 2 000. Raw data (FASTQ files format) from both protocols were filtered by removing adapter sequences and low quality reads using Trimmomatic 0.36 and 0.38 (Bolger et al., 2014) and overlapping reads from protocol 2 were merged using FLASH (Magoc and Salzberg, 2011). The resulting filtered RAW reads from both protocols were submitted to NCBI (Sequence Read Archives (SRA): *pending*). Filtered reads were *de novo* assembled using Trinity v2.2.0 (Protocol 1) and v2.0.6 (Protocol 2) (Grabherr et al., 2011, Haas et al., 2013) with read normalization option. All Trinity transcriptome assemblies used in this study have been submitted to NCBI Transcriptome Shotgun Assembly (TSA) database (pending).

2.3 Compiling of precursor sequences

For the identification of precursor sequences we performed a search by homology (tBLASTn) using as reference queries precursor sequences of polyneopteran insects such as *L. migratoria*, *P. americana* and *C. morosus* (Liessem et al., 2018, Veenstra, 2014, Zeng et al., 2021) on a local computer as implement in BLAST+ (Camacho et al., 2009). Positive hits within the transcriptomes and draft genome assemblies were translated into proteins using ExPASy translate tool (Artimo et al., 2012) (http://web.expasy.org/translate/). Signal peptides for

the putative precursors were predicted using the SignalP 5.0 server (Nielsen, 2017) (www.cbs.dtu.dk/services/SignalP/). Cleavage sites were preliminary assigned based on known cleavage sites in homologous precursors from other species. In addition, missing neuropeptide precursors were also searched in the raw data using the BLAST+ algorithm.

2.4 Tissue Preparation for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)

Animals were kept at -10° C for 5 min before dissecting the nervous system in insect saline (NaCl 126 mM, KCl 5.4 mM, NaH_2PO4 0.17 mM, KH_2PO_4 0.22mM, pH 7.4; Rubakhin and Sweedler, 2007). For direct tissue profiling of peptides by means of MALDI TOF MS (Wegener et al., 2010), we dissected neurohemal organs, including the retrocerebral complex (RCC) and perisympathetic organs (PSOs), neuropile regions such as antennal lobe (AL) and terminal ganglion (TermG), and the frontal ganglion (FG) as part of the stomatogastric system. Dissected preparations were washed for few seconds in a drop of distilled water and then transferred onto a sample plate for MALDI TOF MS. After drying of the preparations on the sample plate, they were covered with 0.3–0.4 µl of matrix solution.

2.5 Tissue preparation for Quadrupole-Orbitrap mass spectrometry (Orbitrap MS)

The nervous system of an adult female and a 5th instar juvenile were dissected in insect saline, washed in a drop of distilled water to remove salt contaminations and transferred into 0.5 ml Safe-Lock tubes (Eppendorf, Hamburg, Germany) containing 30 µl extraction solution (90% methanol, 1% formic acid [FA]). Tissues were disintegrated first in an ultrasonic bath (Transonic 660/H, Elma Schmidbauer GmbH, Hechingen, Germany) at 4° C for 5 min and subsequently with an ultrasonic-probe three times for 3 s (Bandelin Sonopuls HD 200, Bandelin electronic GmbH, Berlin, Germany), respectively. Extracts were then centrifuged for 20 min at 15 000 rpm at 4 °C. Supernatants were transferred in a fresh tube (Eppendorf, Hamburg, Germany) and methanol was evaporated in a vacuum concentrator (Hetovac VR-1, Heto Lab Equipment, Roskilde, Denmark).

2.6 Orbitrap MS

Samples were desalted using self-packed Stage Tip SDB-RPS columns (IVA Analysentechnik e. K., Meerbusch, Germany) spin columns (Rappsilber et al., 2007). Subsequently, peptides were separated on an EASY nanoLC1000 UPLC system (Thermo Fisher Scientific) using in-house packed RPC18-columns 50 cm (fused Silica tube with ID 50 µm ± 3 µm, OD 150 µm ± 6 µm, Reprosil 1.9 µm, pore diameter 60 Å; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and a binary buffer system (A: 0.1% FA, B: 80% ACN, 0.1% FA). Running conditions were as follows: linear gradient from 2 to 62% B in 110 min, 62 to 75% B in 30 min, and final washing from 75 to 95% B in 6 min (45 °C, flow rate 250 nL/min). Finally, the columns were re-equilibrated for 4 min at 5% B. The UPLC was coupled to a Q-Exactive Plus (Thermo Fisher Scientific) mass spectrometer. MS data were acquired in a top 10 data-dependent method dynamically choosing the most abundant peptide ions from the respective survey scans in a range of m/z 300−3 000 for HCD fragmentation. Full MS^1 acquisitions ran with 70 000 resolution, automatic gain control target (AGC target) at 3×10^6 and maximum injection time at 80 ms. HCD spectra were measured with a resolution of 35 000, AGC target at 1×10^6, maximum injection time at 120 ms, 28 eV normalized collision energy, and dynamic exclusion set at 25 s. The instrument was run with peptide recognition mode (2−8 charges), singly charged and unassigned precursor ions were excluded.

Raw data of experiments were analysed with PEAKS Studio 10 (Bioinformatics Solutions Inc., Waterloo, Canada; Zhang et al., 2012). Peptides were searched with a parent error mass tolerance of 10 ppm and fragment

mass error tolerance of 0.05 Da against an internal database containing genome- and transcriptome-derived precursor sequences of *S. gregaria* as well as six frames translation of the transcriptome datasets; none enzyme mode was selected. Variable post-translational modifications included in the searches were: N-terminal acetylation, C-terminal amidation, disulfide bridges, formylation, oxidation at methionine, pyroglutamate from glutamine, pyroglutamate from glutamic acid, and tyrosine sulfation. The false discovery rate (FDR) was determined by the decoy database search implement in PEAKS 10 and set below 1%. To provide the accurate monoisotopic mass of a peptide, Orbitrap MS RAW data were corrected prior to the analysis (precursor mass correction only). Fragment spectra with a peptide score (−10 lgP) equivalent to a P-value of about 1%, were manually reviewed. Peptide spectrum matches with a FDR of 0.1 % (approximately -10logP values higher than 30) were also manually checked.

2.7 MALDI-TOF MS

Mass fingerprints (MS^1) and ion fragmentation spectra (LIFT mode - MS^2) spectra were acquired using either the ultrafleXtreme mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) or the ABI 4800 Proteomics Analyzer (Applied Biosystems Framingham, MA). The ultrafleXtreme was mostly used for MS^1 under manual control in reflectron-positive ion mode and for MS^2 experiments with LIFT technology without CID with both 2,5-dihydroxybenzoic acid (DHB, Sigma-Aldrich) and α-cyano-4-hydroxycinnamic acid (α-CHCA, Sigma-Aldrich) as matrices, while the ABI 4800 syetem was used mainly for MS^2 in gas off mode using α-CHCA. MS^2 spectra were manually reviewed by comparison of experimentally obtained fragment ions with theoretical fragment ion (http://prospector.ucsf.edu). The DHB matrix was prepared by dissolving DHB to a final concentration of 10 mg/ml in 1% aqueous FA containing 20% acetonitrile (ACN) (v/v) and the α-CHCA matrix was prepared with 10 mg/ml CHCA diluted in 60% ethanol, 36% ACN, 4% water (stock solution) and dissolved in 50% methanol/water (2:1 50% methanol/CHCA stock solution) before use and applied to the dry samples. For external calibration we used a mixture of proctolin, *D. melanogaster* short neuropeptide 2^{12-19} (sNPF-2^{12-19}), *P. americana* (Pea)-FMRFa-12, *L. migratoria* periviscerokinin (Lom-PVK), Pea-SKN, and glucagon for the lower mass range (m/z 600–4 000) and a mixture of bovine insulin, glucagon and ubiquitin for the higher mass range (m/z 3 000–10 000). Spectra were analysed using flexAnalysis 3.4 software package (Bruker Daltonik) and Data Explorer v. 4.3 (Applied Biosystems).

3. Results

3.1 *Compilation of precursors based on genome and transcriptome information*

Orthology-based searches in the *S. gregaria* genome database (Verlinden et al., 2020) and transcriptomes using precursor gene information from *L. migratoria* (Hou et al., 2015, Veenstra, 2014, 2021, Veenstra and Šimo, 2020, Veenstra et al., 2021), *C. morosus* (Liessem et al., 2018), and *P. americana* (Zeng et al., 2021) resulted in the identification of precursors of 67 genes encoding neuropeptides and neuropeptide-like sequences (Supplementary Material 1; see Table 1 for full names and abbreviations for precursors). The orthologous genes for many of these precursors were already known from *L. migratoria* (Hou et al., 2015, Veenstra, 2014, EFLa: Veenstra and Šimo, 2020, Gonadulin and IGF: Veenstra et al., 2021, PKL: Redeker et al., 2017, PNPLP/ PaOGS36577: Zeng et al., 2021, iPTH: Xie et al., 2020, hanSolin and RFLa: Bläser and Predel, 2020). However, some of the precursor sequences found in *S. gregaria* were not yet known for *L. migratoria*. These precursor sequences of *L. migratoria* (ALP, AstC, CNPLP, IDL, ITG) are provided in Supplementary Material 2, which also lists all precursor sequences reported since the last comprehensive overview for *L. migratoria* (Veenstra, 2014). Eleven of the 67 precursor genes listed for *S. gregaria* show alternative transcripts (2 transcripts: CAPA, CCHa2, ITP, MIP, MS, NPF1, NVP, OK, sNPF; 3 transcripts: ALP; 4 transcripts: neuroparsin), bringing the total number of *S. gregaria* precursors identified to 81 (Table 1, Supplementary Material 1). The gene structure of a number of alternatively spliced genes is provided in Supplementary Material 3. Most of these precursor sequences are complete (Table 1). Exceptions are the precursors of bursicon alpha, EFLa, and PK, for which the signal peptide sequences are still incomplete, and the precursor of SGSSP, for which the sequence of the C-terminal motif probably lacks in part. Allele-specific differences at the amino acid level were found for the precursors of AstC, calcitonin B, CAPA, CNPLP, EH2, and ITG (Table 1, Supplementary Material 1).

3.2 *Peptidomics of the nervous system: sequence confirmation*

Analyses of tissue extracts from various parts of the central nervous system (CNS) and RCC by Orbitrap MS confirmed by MS^2 experiments the presence of products of 46 of the 67 *S. gregaria* precursors (Table 2). In addition, MALDI-TOF MS^2 experiments verified the presence of ACP, which could not be clearly identified by Orbitrap MS, in CNS extracts (Fig. 1). Peptidomics also confirmed products of alternatively spliced genes for precursors of CCHa2, MS, NPF1, and CAPA, whereas for genes coding MIP, OK, sNPF, NVP, and ITP, products of only one of the two predicted transcripts were detected in the nervous system (Table 2). For ALP, confirmed peptides could not be assigned to a specific splice variant because the corresponding sequences are present in all gene transcripts (see Supplementary Material 1). The extent of precursor sequences confirmed by MS^2 spectra varies considerably for the different precursors. It can be assumed that precursor sequences with complete or near-complete sequence confirmation are highly abundant in the CNS or corpora cardiaca (CC). Among the precursors exhibiting high sequence confirmation by MS^2 spectra are precursors of AKH1, AKH2, AstA, AstCCC (AstC of most previous publications, but see Veenstra, 2016), AT, $CAPA_{a+b}$, CCHa1, CCHa2, Crz, CRF-DH, FMRFa, kinin, MIP, MS_{a+b}, $NPF1_{a+b}$, NPLP1, OK_a, PNPLP, PK, PKL, sNPF, and TKRP. With the exception of the neuropeptide-like precursors NPLP1 and PNPLP, these are neuropeptide precursors of functionally more or less well studied neuropeptides. The information obtained in this way about identified sequences can be combined with information on relative signal intensities of the peptides in MALDI TOF mass spectra (see below) to obtain an even better overall view. For the above mentioned precursors, Orbitrap MS^2 spectra also always confirmed the predicted mature bioactive peptides, including PTMs such as pyroglutamate

formation and C-terminal amidation (Table 2). Similarly, mature peptides such as ACP (MALDI-TOF MS^2), ALP, CT-DH, ETH, hanSolin, Nat, PDF, proctolin, SIFa, SKs, and tryptoPKs were also confirmed (Table 2), but sequence confirmation over the entire precursor was less complete for the respective precursors. For several abundant neuropeptides (AKH1, AstCCC, AT, PVK-1 of CAPA) whose position in the precursor is directly C-terminal to the signal peptide, multiple truncated and/or extended N-termini were found, suggesting variable C-terminal cleavages of the signal peptides. Intermediate products of precursor processing were found for AKH1, AKH2, and the two CAPA-tryptoPKs (Table 2). In these sequences, the C-termini were not yet fully processed. The products of the following precursors could not be detected in samples of the nervous system or RCC: AKH3, AstC, bursicon alpha/beta, calcitonin A and B, CNMa, EFLa, gonadulin, elevenin, inotocin, IGF, iPTH, SGSSP, SMYa, trissin, EH, GPA2, GPB5, and neuroparsin.

3.3 *Peptidomics of the nervous system: tissue-specific distribution of peptides*

Direct tissue profiling by MALDI-TOF MS was used by us to examine the tissue-specific distribution of abundant peptides in the nervous system. For this purpose, we analyzed two neuropil regions within the CNS known to be innervated by a multitude of peptidergic cells (antennal lobe/AL + posterior neuropil of the terminal ganglion/TermG), furthermore the frontal ganglion/FG as part of the stomatogastric nervous system as well as the major release sites of peptide hormones from the CNS (CC, abdominal + thoracic perisympathetic organs/PSOs). The corresponding mass spectra are MS^1 spectra, meaning that ion signals with mass identity to *S. gregaria* peptides are labeled with the respective peptide names, although these substances were mostly not fragmented from these samples. Sequence confirmations of these peptides, either by MALDI-TOF MS^2 experiments or Orbitrap MS, are listed in Table 2.

Antennal lobe (Fig. 2A): This neuropil contains numerous neuropeptides and neuropeptide-like substances, and only the peptides with higher signal intensity are labeled in Fig. 2A. Many of the prominent ion signals, including the highest signal (NVP), are mass-identical to peptides from neuropeptide-like precursors (NVP, NPLP1, PNPLP). Ion signals that are mass-identical to peptides of genuine neuropeptide precursors include AstA, kinin, sNPF, TKRP, MIP, MS_{a+b}, AT, OK_a, SIFa and/or AstCCC (mass similarity), CCHa2, NPF1, and CT-DH. In addition, a very prominent ion signal with mass identity to the C-terminal precursor peptide of ITG is detectable. Of the two MS sequences resulting from alternative transcripts, MS_a shows significantly higher signal intensity. This was generally observed in all spectra containing ion signals typical of MS. Of the two (nearly) mass-identical SIFa/AstCCC, it is known for SIFa that this peptide is present in ALs of *S. gregaria* (Gellerer et al., 2015); both peptides were biochemically confirmed in brain extracts (Gellerer et al., 2015).

Terminal ganglion (Fig. 2B): Mass spectra of the caudal neuropil of the TermG support the presence of all peptides, including those from neuropeptide-like precursors, identified in the AL, although the relative abundances are not identical (e.g., more MIP, less kinin and NPF1-PP). In addition, prominent ion signals with mass identity to ACP and AstCC are detectable.

Frontal ganglion (Fig. 2C): Preparations from the FG, which links the stomatogastric nervous system to the cerebral ganglia, revealed a peptidome quite similar to that of selected parts of the CNS (see Fig. 2A, B). Prominent ion signals typical of kinins or AstCC, were not detectable. The most abundant ion signals indicate enrichment of MS and sNPF. In addition, ion signals mass-identical to SKs are present. Due to the considerable signal intensity of the ion signal at 1650.7 (SIFa and/or AstCCC), the corresponding peptide(s) could be

fragmented using the same preparation and fragment analysis (not shown) verified the presence of AstCCC in the FG.

Corpora cardiaca (Fig. 2D): The CC of insects consists of a glandular part (synthesis of, e.g., AKH) and a neurohemal part that stores peptide hormones of the cerebral and gnathal ganglia. The mass spectrum shown in Fig. 2D represents a preparation of the glandular part with adjacent neurohemal tissue. AKH1 and AKH2 are mainly detectable with their sodium adduct ions. The ion signals of the very abundant AKH1 are accompanied by ion signals with mass identity to the monomer as well as to the dimer of APRP of the corresponding precursor. The identity of the APRP1 monomer was verified by MS^2 experiments using the same sample (not shown). Other prominent ion signals represent neuropeptides synthesized in the CNS, including AstA, PK, PKL, sNPF, ACP, MS_a, and a precursor peptide of the CRF-DH precursor (Fig. 2D), whereas relatively weak ion signals indicate the presence of Crz, hanSolin, MS_b, and CCHa2. Ion signals with mass identity to peptides of the neuropeptide-like precursors were either not observed or detected with very low signal intensity compared with the CNS samples. The presence of the IRP is supported by ion signals typical of the beta chain of IRP and two IRP precursor-derived peptides.

Thoracic perisympathetic organs (Fig. 2E): Mass spectra from thoracic PSO preparations are dominated by FMRFa. The three short FMRFa (FMRFa-2, 3, 4) show by far the highest signal intensity, while the longer FMRFa (FMRFa-1, 5) are not represented by clearly defined single mature peptides. Instead, these FMRFa occur with N-termini of variable lengths. In addition to peptides from the FMRFa precursor, thoracic PSO preparations show the presence of peptides from the NPF1 precursor.

Abdominal perisympathetic organs (Fig. 2F): Mass spectra from abdominal PSO preparations show exclusively CAPA peptides. Among CAPA-PVKs, mature PVK-2 and 4 are highly enriched, while PVK-1 occurs with N-termini of varying lengths. The most abundant PVK-1 peptide, referred to here as mature PVK-1, is cleaved immediately C-terminally of the predicted signal peptide. This peptide has a potential internal cleavage site (Lys-Lys) that is only modestly used (Fig. 2F). The resulting truncated PVK-1 is identical to a peptide described as PKL-1 (Veelaert et al., 1997). The predicted mature PVK-3 is present only at low abundance, but longer sequences that include N-terminally CAPA-PP-2 are detectable. Of the two CAPA-tryptoPKs, tryptoPK-1 is particularly highly enriched. The less abundant tryptoPK-2 is accompanied by a C-terminally extended form that is cleaved at an Arg-Arg cleavage site downstream of the primary cleavage motif. The low relative ion intensities of mature tryptoPK-2 and its extended form, compared with those of tryptoPK-1, suggest incomplete processing of a C-terminally much longer motif containing the CAPA-PK sequence and the C-terminal PP. However, this hypothetical long peptide is above the mass range analyzed. The CAPA-PK was detectable in trace amounts only in few mass spectra.

In addition to the above-mentioned tissues from the central or stomatogastric nervous system, we analyzed by direct tissue profiling the lateral cardiac nerve and tracheal trunks of the abdomen (late-instar nymphs). Mass spectra of the lateral cardiac nerve which is potentially innervated by tryptoPK expressing neurons of the gnathal ganglia (see Redeker et al. 2017) show a large number of ion signals that are mass-identical to tryptoPKs of the tryptoPK precursors (Fig. 3A). Mass spectra of tracheal trunks, the latter presumably containing ETH-expressing Inka cells (Lenaerts et al., 2017), show ion signals of the mature ETHs (Fig. 3B). The identity of these peptides was confirmed by MS^2 using the same samples (not shown).

4. Discussion

Our analysis of neuropeptides and neuropeptide-like substances of *S. gregaria* yielded the most comprehensive peptidomics overview from a locust to date. This is not really surprising. Updated compilations naturally add new information to existing data and take advantage of latest technologies. However, Orbitrap MS data in particular should always be interpreted with some caution. As confirmed sequences become shorter, more false positives occur, and if such data are not supported by verification of predicted mature peptides within the same dataset, the data are not of high significance. Therefore, a simple and fast method such as direct tissue profiling by MALDI-TOF MS often provides important background information on the relative abundance of neuropeptides in specific tissues (see Predel, 2001), thus also avoiding overinterpretation of artificially truncated peptides or those occurring naturally but having a very low abundance. In the end, the more enriched mature neuropeptides derived from the various precursors are usually the physiologically more important peptides. Therefore, it is primarily not the total number of peptides that is important, but an overview of those peptides that are abundant in specific regions of the nervous system.

Among the many precursors described here for the first time for *S. gregaria* are several which have not previously been listed for *L. migratoria* either, although genomes and transcriptomes of the latter species have already been thoroughly screened with respect to neuropeptide precursors (Hou et al., 2015, Veenstra, 2014). This can be explained mainly by the very recent descriptions of some precursors that were previously unknown in insects (Sturm et al., 2016: ALP, Liessem et al., 2018: RFLa, hanSolin, CNPLP, Redeker et al., 2017: PKL, Veenstra et al., 2021: IGF, gonadulin, Xie et al., 2020: iPTH, Zeng et al., 2021: PNPLP). Two of these novel precursors, the iPTH and PKL precursors, contain neuropeptides that almost certainly activate neuropeptide receptors in *S. gregaria*. While we did not find iPTH peptides in the nervous system of *S. gregaria*, PKLs are accumulated in the RCC, similar to sequence-related PKs. One of the PKLs was previously described as PK-6 from this species (Clynen et al., 2003b). Peptides from the remaining novel precursors were also detected in the nervous system. It is less likely that the multiple and not sequence-related peptides from neuropeptide-like precursors such as PNPLP and CNPLP, but also ALP, activate specific neuropeptide receptors. This may also be true for peptides of the NVP precursor, which has been described from *Apis mellifera* (Hummon et al., 2006). However, the peptides of neuropeptide-like precursors are easily detectable in the CNS and therefore represent an integral and prominent part of the peptidome of the nervous system. Whatever their function, neuropeptide-like molecules always contribute to the peptidome of the CNS, and detailed knowledge of their mature products generally also supports interpretation of the genuine neuropeptidome. An interesting aspect in this context is the apparent accumulation of products of neuropeptide-like precursors within the ganglia of the nervous system, whereas only trace amounts of these peptides were found in neurohemal release sites (CC, tPSO, aPSO). Obviously, neuropil regions within the insect nervous system not only show an enrichment of neuropeptide-like substances, but also contain an impressive number of true neuropeptides. This was already known for ALs of various insects (Binzer et al., 2014, Neupert et al., 2011, Siju et al., 2014) and especially for the caudal neuropil region of the TermG (Predel et al., 2010). Here we used tissue samples from these neuropil regions of the CNS as well as from the FG primarily to provide an overview of the mature peptides present in the CNS. Neuropeptides of the following precursors appear to be generally abundant in the nervous system of *S. gregaria* and are not restricted to any specific neuropil: AstA, sNPF, TKRP, MIP, MS, AT, OK_a, CCHa2, NPF1, and CT-DH. Others such as kinins, ACP, AstCC, and probably the similar-mass SIFa and AstCCC were less regularly

present in our mass spectra from samples of the nervous system, but all mature neuropeptides of the precursors mentioned here were confirmed not only by mass match in MALDI-TOF mass spectra but also by additional MS^2 data. There were similar sequence confirmations for peptide hormones produced in the gnathal ganglia (products of PK, PKL, and tryptoPK precursors), thoracic ganglia (FMRFa), and abdominal ganglia (products of CAPA precursor); these hormones are particularly accumulated in the respective neurohemal release sites.

Owing to the abundance of neuropeptides in neurohemal tissues, direct tissue profiling of these hormone release sites often provides an unmatched overview of major products of a neuropeptide gene. In addition, differential expression of duplicated genes, the presence of different splice variants or alleles, and also differential processing of precursors are readily detectable when the corresponding products are accumulated in neurohemal tissues. A good example of this is the sequence-related PKs, PKLs, tryptoPKs, and PVKs. There is no other group of neuropeptides/peptide hormones in insects that is so regularly misinterpreted. Different names for related peptides or even orthologous genes (e.g., *pyrokinin/pk* gene named after the first fully sequenced FXPRLamide in Holman et al. (1986) = *pheromone biosynthesis-activating neuropeptide/pban* (Kawano et al., 1992, Raina et al., 1989) = *hugin* (Meng et al., 2002) are only part of the problem. Since we had to rename some peptides for *S. gregaria* as well, we give here some more detailed information/hypotheses on the evolution of these genes or their peptides, focusing on the information relevant to the situation in locusts. 1) The most recent common ancestor of Hexapoda possessed a *capa/pk* gene encoding in this order PVKs, tryptoPK, PKs (Derst et al., 2016). Specific receptors for each of these sequence-related peptides are known, at least in *D. melanogaster* (Predel and Wegener, 2006). We speculate that different receptors for PK, tryptoPK, and PVK are also present in locusts. 2) Winged insects (Pterygota) have separate *pk* and *capa* genes. In very simplified form, it can be stated that the CAPA precursor represents the N-terminus (originally multiple PVKs + tryptoPK + a single PK) and the PK precursor the C-terminus (tryptoPK + multiple PKs) of the ancient precursor. Possibly, the different expression pattern of the novel *capa* and *pk* genes, which is also typical of *S. gregaria*, was already typical of the most recent ancestor of Pterygota + Zygentoma/silverfish (Diesner et al., 2021). 3) While large insect groups such as Holometabola and Hemiptera appear to use differential processing to release only tryptoPKs of the *capa* gene but not PVKs from CAPA neurons in the labial neuromere of gnathal ganglia (Neupert et al., 2009), novel genes encoding exclusively tryptoPKs emerged in Polyneoptera (Veenstra, 2014). In locusts, these *tryptopk* genes are expressed exclusively in those neurons of the labial neuromere (Redeker et al., 2017) that are homologous to the aforementioned CAPA neurons in the gnathal ganglia of Holometabola and Hemiptera. These neurons have particular projections within the CNS of locusts, including projecting into the lateral cardiac nerve via the ventral nerve cord (Bräunig, 1991). We used this knowledge to confirm by MALDI-TOF MS that *tryptopk* expressing neurons of *S. gregaria* contain neither PK nor PVK. 4). In addition to the *capa*, *tryptopk*, and *pk* genes, locusts possess a *pkl* gene (Redeker et al., 2017) that is also found in *S. gregaria* and appears to have an identical expression pattern to the *pk* gene. On the evolutionary time scale, the *pkl* gene is the most recent addition and codes only for PKs.

Previous transcriptome and genome data indicated alternative splicing for a number of genes in *L. migratoria*, including *itp*, *ms*, *neuroparsin*, *npf1*, and *ok* (Veenstra 2014, Hou et al. 2015). Of the resulting precursors, the two OK precursors do not share any identical neuropeptide. In insects, these OK transcripts are tissue-specific (Sterkel et al., 2012) which explains why we detected, by mass spectrometry, only neuropeptides derived from precursor OK_a (Homberg et al., 2021). The remaining transcripts are short and long forms of precursors that

otherwise contain mostly identical neuropeptides. Our transcriptome and genome data on *S. gregaria* suggest alternative splicing for all these genes and, in addition, for *alp, capa, ccha2, mip, nvp,* and *snpf*. Alternative transcripts of NPF1 leading to a longer/shorter form of NPF (not be confused with sNPF, see Nässel and Wegener, 2011) are a basic feature of Hexapoda (Derst et al., 2016) and are also typical of other Pancrustacea (e.g., *Daphnia*, Dircksen et al., 2011). On the other hand, alternative splicing of CCHa2 seems to be common only in Caelifera (grasshoppers), and alternative transcripts resulting in slightly different MS sequences are only known from few Caelifera (incl. *L. migratoria, S. gregaria*) and a single Dermapteran (Bläser and Predel, 2020). Our peptidomics data confirmed in CNS samples mature peptides from two predicted transcripts of *ccha2*, *ms*, *npf*1, and also the *capa* gene.

Detailed evaluation of the MALDI-TOF mass spectra from various nervous and neuroendocrine tissues (Fig. 2) confirmed a remarkably complete assignment of prominent ion signals to predicted and biochemically confirmed sequences of neuropeptides and neuropeptide-like substances. Compared to previous attempts that used direct tissue profiling of different tissues in *S. gregaria* (Clynen and Schoofs, 2009), the largest fraction which was completely missing in the earlier studies is that of neuropeptide-like substances. In addition, the assignment of an ion signal at 1591.7 to AG-MT-2 (accessory gland myotropin; Clynen et al. 2009) is rather uncertain. It is likely that this ion signal represents the mass-identical CCHa2; at least, we did not obtain any sequence confirmation for the presence of AG-MT-2 in the CNS of *S. gregaria*. For Locusta AG-MT-2, it has generally been questioned (Veenstra 2014) whether it is a neuropeptide because of the lack of upstream and downstream convertase cleavage sites in the precursor. Genes whose products could not be detected in our or previous mass spectrometry-based analyses include, in particular, those too large to be detected by the peptidomic approaches employed here (bursicon, EH, gonadulin, GPA2, GPB5, IGF, neuroparsin) or neuropeptide genes whose mature peptides have never or rarely been detected biochemically in the nervous system of insects (AstC, calcitonin, CNMa, EFLa, elevenin, inotocin, SMYa, trissin). Some of the latter neuropeptides may be synthesized in only a few neurons and could be detected in the CNS by targeted single cell mass spectrometry (see, e.g., *C. morosus* elevenin; Liessem et al. 2018).

5. Conclusion

The dataset for *S. gregaria* presented here confirms substantial conservation of the entire gene set for neuropeptides and neuropeptide-like peptides typical of insects. This conservation is also found in other Polyneoptera that additionally have novel neuropeptide genes resulting from gene duplications (*smya*, *tryptopk*), and this contrasts strikingly with substantial losses of peptidergic systems in holometabolous species such as *D. melanogaster* (Bläser and Predel 2020). Apart from species-specific sequences within the known range of variation, the neuropeptidome of *S. gregaria* is quite similar to that of *L. migratoria* and even to that of less closely related Polyneoptera. Therefore, our peptidomics data on the mature peptides derived from the various precursors constitute a very useful resource that may prove to be of more general relevance, even when considering other species of Polyneoptera.

ACKNOWLEDGEMENTS

The authors would like to thank Evelien Herinckx for taking care of the locust rearing facility at KU Leuven. We thank Christian Frese, Astrid Wilbrand-Hennes, Jan Krueger and Ursula Cullman (CECAD Cologne Proteomics

Facility) for the Orbitrap MS analyses, Tobias Schulz (Biocenter Cologne) for IT support, and Marek Franitza and Janine Altmüller for transcriptome sequencing at Cologne Center for Genomics. This work was supported by grants from the Special Research Fund of KU Leuven [grant numbers C14/15/050, C14/19/069], the European Union's Horizon 2020 Research and Innovation programme [grant number 634361 to RP and to JVB], and the Research Foundation of Flanders (FWO) (G0F2417N, G090919N) to JVB.

References

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research 40, W597–W603. https://doi.org/10.1093/nar/gks400

Badisco, L., Claeys, I., Van Hiel, M., Clynen, E., Huybrechts, J., Vandersmissen, T., Van Soest, S., Vanden Bosch, L., Simonet, G., Vanden Broeck, J., 2008. Purification and characterization of an insulin-related peptide in the desert locust, *Schistocerca gregaria*: immunolocalization, cDNA cloning, transcript profiling and interaction with neuroparsin. Journal of Molecular Endocrinology 40, 137–150. doi: 10.1677/jme-07-0161

Badisco, L., Huybrechts, J., Simonet, G., Verlinden, H., Marchal, E., Huybrechts, R., Schoofs, L., de Loof, A., Vanden Broeck, J., 2011. Transcriptome analysis of the desert locust central nervous system: Production and annotation of a *Schistocerca gregaria* EST database. PLoS ONE 6, e17274. https://doi.org/10.1371/journal.pone.0017274

Binzer, M., Heuer, C.M., Kollmann, M., Kahnt, J., Hauser, F., Grimmelikhuijzen, C.J.P., Schachtner, J., 2014. Neuropeptidome of *Tribolium castaneum* antennal lobes and mushroom bodies. Journal of Comparative Neurology 522, 337–357. https://doi.org/10.1002/cne.23399

Bläser, M., Predel, R., 2020. Evolution of neuropeptide precursors in polyneoptera (Insecta). Frontiers in Endocrinology 11, 197. https://doi.org/10.3389/fendo.2020.00197

Bräunig, P., 1991. A suboesophageal ganglion cell innervates heart and retrocerebral glandular complex in the Locust. Journal of Experimental Biology 156, 567–582. https://doi.org/10.1242/jeb.156.1.567

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Chang, J., Zhao, J., Tian, X., 2018. In silico prediction of neuropeptides in Hymenoptera parasitoid wasps. PLoS ONE 13, e0193561. https://doi.org/10.1371/journal.pone.0193561

Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z., Kang, L., 2010. De Novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. PLoS ONE 5, e15633. https://doi.org/10.1371/journal.pone.0015633

Clynen, E., Baggerman, G., Veelaert, D., Cerstiaens, A., Van Der Horst, D., Harthoorn, L., Derua, R., Waelkens, E., De Loof, A., Schoofs, L., 2001. Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, *Locusta migratoria*. European Journal of Biochemistry 268, 1929–1939. https://doi.org/10.1046/j.1432-1327.2001.02067.x

Clynen, E., Huybrechts, J., De Loof, A., Schoofs, L., 2003a. Mass spectrometric analysis of the perisympathetic organs in locusts: identification of novel periviscerokinins.Biochemical and biophysical research communications 300,2: 422-8. doi:10.1016/s0006-291x(02)02869-3

Clynen, E., Baggerman, G., Huybrechts, J., Vanden Bosch, L., De Loof, A., Schoofs, L., 2003b. Peptidomics of the locust *corpora allata*: Identification of novel pyrokinins (-FXPRLamides). Peptides 24, 1493–1500. https://doi.org/10.1016/j.peptides.2003.10.006

Clynen, E., Huybrechts, J., Baggerman, G., Van Doorn, J., Van der Horst, D., De Loof, A., Schoofs, L., 2003c.

Identification of a Glycogenolysis-Inhibiting Peptide from the Corpora Cardiaca of Locusts, Endocrinology 144(8), 3441–3448. https://doi.org/10.1210/en.2002-0107

Clynen, E., Schoofs, L., 2009. Peptidomic survey of the locust neuroendocrine system. Insect Biochemistry and Molecular Biology 39, 491–507. https://doi.org/10.1016/j.ibmb.2009.06.001

Clynen, E., Husson, S.J., Schoofs, L., 2009. Identification of new members of the (short) neuropeptide F family in locusts and *Caenorhabditis elegans*. Annals of the New York Academy of Sciences 1163:60-74. doi: 10.1111/j.1749-6632.2008.03624.x.)

Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., Predel, R., 2016. Evolution of neuropeptides in non-pterygote hexapods. BMC evolutionary biology 16, 1–10. https://doi.org/10.1186/s12862-016-0621-4

Diesner, M., Bläser, M., Eckardt, S., Iliffe, T.M., Boelen Theile, E. And Predel R. 2021. Expression pattern of CAPA/pyrokinin neuropeptide genes in Remipedia and silverfish: Rapid differentiation after gene duplication in early Hexapoda, followed by strong conservation of newly established features in insects. Peptides Volume 144, 170610. https://doi.org/10.1016/j.peptides.2021.170610

Dillen, S., Verdonck, R., Zels, S., Van Wielendaele, P., Vanden Broeck, J., 2014. Identification of the short neuropeptide F precursor in the desert locust: evidence for an inhibitory role of sNPF in the control of feeding. Peptides 53, 134-139. doi: 10.1016/j.peptides.2013.09.018

Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J.P., 2011. Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein hormones. Journal of Proteome Research 10, 4478–4504. https://doi.org/10.1021/pr200284e

Elphick, M.R., Mirabeau, O., Larhammar, D., 2018. Evolution of neuropeptide signalling systems. Journal of Experimental Biology 221. https://doi.org/10.1242/jeb.151092

Gäde, G., Goldsworthy, G.J., Schaffer, M.H., Cook, J.C., Rinehart K.L., Jr, 1986. Sequence analyses of adipokinetic hormones II from *corpora cardiaca* of *Schistocerca nitans*, *Schistocerca gregaria*, and *Locusta migratoria* by fast atom bombardment mass spectrometry. Biochemical and Biophysical Research Communications 134(2), 723–730. doi:10.1016/S0006-291X(86)80480-6.

Gellerer, A., Franke, A., Neupert, S., Predel, R., Zhou, X., Liu, S., Reiher, W., Wegener, C., Homberg, U., 2015. Identification and distribution of sifamide in the nervous system of the desert locust *Schistocerca gregaria*. Journal of Comparative Neurology 523, 108–125. https://doi.org/10.1002/cne.23671

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29, 644–652. https://doi.org/10.1038/nbt.1883

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hekimi, S., Burkhart, W., Moyer, M., Fowler, E., O'Shea M., 1989. Dimer structure of a neuropeptide precursor established: consequences for processing. Neuron 2(4), 1363-8. doi:10.1016/0896-6273(89)90074-3

Hofer, S., Dircksen, H., Tollbäck, P., Homberg U., 2005. Novel insect orcokinins: characterization and neuronal distribution in the brains of selected dicondylian insects. The Journal of Comparative Neurology 490(1), 57-71. doi:10.1002/cne.20650

Holman, G.M., Cook, B.J., Nachman, R.J., 1986. Primary structure and synthesis of a blocked myotropic neuropeptide isolated from the cockroach, *Leucophaea maderae*. Comparative Biochemistry and Physiology Part C, Comparative 85, 219–224. https://doi.org/10.1016/0742-8413(86)90077-0

Holman, G.M., Nachman, R.J., Wright, M.S., Schoofs, L., Hayes, T.K., DeLoof, A., 1991. Insect myotrophic

peptides isolation, structural characterization, and biological activities. ACS symposium series American Chemical Society 40–50. https://doi.org/DOI: 10.1021/bk-1991-0453.ch005

Homberg, U., Hensgen, R., Rieber, E., Seyfarth, J., Kern, M., Dippel, S., Dircksen, H., Spänig, L., Kina, Y.P., 2021. Orcokinin in the central complex of the locust *Schistocerca gregaria*: Identification of immunostained neurons and colocalization with other neuroactive substances. Journal of Comparative Neurology 529, 1876–1894. https://doi.org/10.1002/cne.25062

Hou, L., Jiang, F., Yang, P., Wang, X., Kang, L., 2015. Molecular characterization and expression profiles of neuropeptide precursors in the migratory locust. Insect Biochemistry and Molecular Biology 63, 63–71. https://doi.org/10.1016/j.ibmb.2015.05.014

Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A., Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J. V., 2006. From the genome to the proteome: Uncovering peptides in the *Apis* brain. Science 314, 647–649. https://doi.org/10.1126/science.1124128

Jékely, G., 2013. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proceedings of the National Academy of Sciences 110, 8702 LP – 8707. https://doi.org/10.1073/pnas.1221833110

Kawano, T., Kataoka, H., Nagasawa, H., Isogai, A., Suzuki, A., 1992. cDNA cloning and sequence determination of the pheromone biosynthesis activating neuropeptide of the silkworm, *Bombyx mori*. Biochemical and Biophysical Research Communications 189, 221–226. https://doi.org/10.1016/0006-291X(92)91547-4

Kopeć, S., 1922. Studies on the necessity of the brain for the inception of insect metamorphosis. The Biological Bulletin 42, 323–342. https://doi.org/10.2307/1536759

Lenaerts, C., Cools, D., Verdonck, R., Verbakel, L., Vanden Broeck, J., Marchal, E., 2017. The ecdysis triggering hormone system is essential for successful moulting of a major hemimetabolous pest insect, *Schistocerca gregaria*. Scientific Reports 7, 46502. https://doi.org/10.1038/srep46502

Liessem, S., Ragionieri, L., Neupert, S., Büschges, A., Predel, R., 2018. Transcriptomic and neuropeptidomic analysis of the stick insect, *Carausius morosus*. Journal of Proteome Research 17, 2192–2204. https://doi.org/10.1021/acs.jproteome.8b00155

Meng, X., Wahlström, G., Immonen, T., Kolmer, M., Tirronen, M., Predel, R., Kalkkinen, N., Heino, T.I., Sariola, H., Roos, C., 2002. The *Drosophila* hugin gene codes for myostimulatory and ecdysis-modifying neuropeptides. Mechanisms of Development 117, 5–13. https://doi.org/10.1016/S0925-4773(02)00175-2

Meredith, J., Ring, M., Macins, A., Marschall, J., Cheng, N.N., Theilmann, D., Brock, H.W., Phillips, J.E., 1996. Locust ion transport peptide (ITP): primary structure, cDNA and expression in a baculovirus system. The Journal of experimental biology 199(5): 1053-1061.

Nässel, D.R., Wegener, C., 2011. A comparative review of short and long neuropeptide F signaling in invertebrates: Any similarities to vertebrate neuropeptide y signaling? Peptides 32, 1335–1355. https://doi.org/10.1016/j.peptides.2011.03.013

Nässel, D.R., Zandawala, M., 2019. Recent advances in neuropeptide signaling in *Drosophila*, from genes to physiology and behavior. Progress in Neurobiology 179, 1–79. https://doi.org/10.1016/j.pneurobio.2019.02.003

Nässel, D.R., Zandawala, M., 2020. Hormonal axes in *Drosophila*: regulation of hormone release and multiplicity of actions. Cell and Tissue Research 382, 233–266. https://doi.org/10.1007/s00441-020-03264-z

Neupert, S., Huetteroth, W., Schachtner, J., Predel, R., 2009. Conservation of the function counts: Homologous neurons express sequence-related neuropeptides that originate from different genes. Journal of Neurochemistry 111, 757–765. https://doi.org/10.1111/j.1471-4159.2009.06361.x

Neupert, S., Fusca, D., Schachtner, J., Kloppenburg, P., Predel, R., 2011. Toward a single-cell-based analysis of neuropeptide expression in *Periplaneta americana* antennal lobe neurons. Journal of Comparative Neurology 520, 694–716. https://doi.org/10.1002/cne.22745

Nielsen, H., 2017. Predicting secretory proteins with SignalP, in: Kihara, D. (Ed.), Protein Function Prediction: Methods and Protocols. Springer New York, New York, NY, pp. 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6

Paemen, L., Tips, A., Schoofs, L., Proost, P., Van Damme, J., De Loof, A., 1991. Lom-AG- myotropin: a novel myotropic peptide from the male accessory glands of *Locusta migratoria*. Peptides. 12(1), 7-10. doi: 10.1016/0196-9781(91)90158-l.

Predel, R., 2001. Peptidergic neurohemal system of an insect: Mass spectrometric morphology. Journal of Comparative Neurology 436, 363–375. https://doi.org/10.1002/cne.1073

Predel, R., Nachman, R.J., Gäde, G., 2001. Myostimulatory neuropeptides in cockroaches: Structures, distribution, pharmacological activities, and mimetic analogs. Journal of Insect Physiology. https://doi.org/10.1016/S0022-1910(00)00129-3

Predel, R., Gäde, G., 2002. Identification of the abundant neuropeptide from abdominal perisympathetic organs of locusts. Peptides 23, 621-628. DOI: 10.1016/S0196-9781(01)00669-6

Predel, R., Wegener, C., 2006. Biology of the CAPA peptides in insects. Cellular and Molecular Life Sciences 63, 2477–2490. https://doi.org/10.1007/s00018-006-6187-3

Predel, R., Neupert, S., Garczynski, S.F., Crim, J.W., Brown, M.R., Russell, W.K., Kahnt, J., Russell, D.H., Nachman, R.J., 2010. Neuropeptidomics of the mosquito *Aedes aegypti*. Journal of Proteome Research 9, 2006–2015. https://doi.org/10.1021/pr901187p

Raina, A.K., Jaffe, H., Kempe, T.G., Keim, P., Blacher, R.W., Fales, H.M., Riley, C.T., Klun, J.A., Ridgway, R.L., Hayes, D.K., 1989. Identification of a neuropeptide hormone that regulates sex pheromone production in female moths. Science (New York, NY) 244, 796–8. https://doi.org/10.1126/science.244.4906.796

Rappsilber, J., Mann, M., Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols 2, 1896–1906. https://doi.org/10.1038/nprot.2007.261

Redeker, J., Bläser, M., Neupert, S., Predel, R., 2017. Identification and distribution of products from novel tryptopyrokinin genes in the locust, *Locusta migratoria*. Biochemical and Biophysical Research Communications 486, 70–75. https://doi.org/https://doi.org/10.1016/j.bbrc.2017.02.135

Robb, S., Packman, L.C., Evans, P.D., 1989. Isolation, primary structure and bioactivity of schistoflrf-amide, a FMRF-amide-like neuropeptide from the locust, *Schistocerca gregaria*. Biochemical and biophysical research communications 160(2), 850-856. doi:10.1016/0006-291x(89)92512-6

Rubakhin, S.S., Sweedler, J. V., 2007. Characterizing peptides in individual mammalian cells using mass spectrometry. Nature Protocols 2, 1987–1997. https://doi.org/10.1038/nprot.2007.277

Santos, D., Vanden Broeck, J., Wynant, N., 2014. Systemic RNA interference in locusts: Reverse genetics and possibilities for locust pest control. Current Opinion in Insect Science 6, 9–14. https://doi.org/10.1016/j.cois.2014.09.013

Scharrer, B., 1987. Insects as Models in Neuroendocrine Research. Annual Review of Entomology 32, 1–16. https://doi.org/10.1146/annurev.en.32.010187.000245

Schoofs, L., Holman, G.M., Hayes, T.K., Nachman, R.J., De Loof, A., 1991. Isolation, identification and synthesis of locustamyoinhibiting peptide (Lom-MIP), a novel biologically active neuropeptide from *Locusta migratoria*. Regulatory Peptides 36, 111–119. DOI: 10.1016/0167-0115(91)90199-q

Schoofs, L., Holman, G.M., Proost, P., Van Damme, J., Hayes, T.K., De Loof, A., 1992. Locustakinin, a novel myotropic peptide from *Locusta migratoria*, isolation, primary structure and synthesis. Regulatory Peptides 37(1), 49-57. doi: 10.1016/0167-0115(92)90063-z

Schoofs, L., Broeck, J. Vanden, De Loof, A., 1993. The myotropic peptides of *Locusta migratoria*: Structures, distribution, functions and receptors. Insect Biochemistry and Molecular Biology 23, 859–881. https://doi.org/10.1016/0965-1748(93)90104-Z

Schoofs, L., Clynen, E., Cerstiaens, A., Baggerman, G., Wei Z., Vercammen, T., Nachman, R., De Loof, A., Tanaka, S., 2001. Newly discovered functions for some myotropic neuropeptides in locusts. Peptides 22, 219–227. doi: 10.1016/S2095-3119(19)62804-2

Schoofs, L., De Loof, A., Van Hiel, M.B., 2017. Neuropeptides as regulators of behavior in insects. Annual Review of Entomology 62, 35–52. https://doi.org/10.1146/annurev-ento-031616-035500

Siegert, K.J., 1999. Locust *corpora cardiaca* contain an inactive adipokinetic hormone. FEBS Letters 447, 237–240. https://doi.org/10.1016/S0014-5793(99)00299-9.

Siju, K.P., Reifenrath, A., Scheiblich, H., Neupert, S., Predel, R., Hansson, B.S., Schachtner, J., Ignell, R., 2014. Neuropeptides in the antennal lobe of the yellow fever mosquito, *Aedes aegypti*. Journal of Comparative Neurology 522, 592–608. https://doi.org/10.1002/cne.23434

Starratt, A.N., Brown, B.E., 1975. Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life Sciences 17, 1253–1256. https://doi.org/10.1016/0024-3205(75)90134-4

Sterkel, M., Oliveira, P.L., Urlaub, H., Hernandez-Martinez, S., Rivera-Pomar, R., Ons, S., 2012. OKB, a novel family of brain-gut neuropeptides from insects. Insect Biochemistry and Molecular Biology 42, 466–473.

Stone, J.V., Mordue, W., Batley, K.E., Morris, H.R., 1976. Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilisation during flight. Nature 263(5574), 207–211. DOI 10.1038/263207a0.

Sturm, S., Ramesh, D., Brockmann, A., Neupert, S., Predel, R., 2016. Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. Journal of Proteomics 132, 77–84. https://doi.org/doi.org/10.1016/j.jprot.2015.11.021

Tawfik, I.A., Tanaka, S., De Loof, A., Schoofs. L., Baggerman, G., Waelkens, E., Derua, R., Milner, Y., Yerushalmi, Y., Pener, M.P., 1999. Identification of the gregarization-associated dark-pigmentotropin in locusts through an albino mutant. Proceedings of National Academy of Science U.S.A. 96, 7083–7087. https://doi.org/10.1073/pnas.96.12.7083

Van Wielendaele, P., Dillen, S., Marchal, E., Badisco, L., Vanden Broeck, J., 2012. CRF-like diuretic hormone negatively affects both feeding and reproduction in the desert locust, *Schistocerca gregaria*. PLoS One 7(2), e31425. doi: 10.1371/journal.pone.0031425

Van Wielendaele, P., Wynant, N., Dillen, S., Badisco, L.,Marchal, E., Vanden Broeck, J., 2013. In vivo effect of Neuropeptide F on ecdysteroidogenesis in adult female desert locusts (*Schistocerca gregaria*). Journal of insect physiology 59(6), 624-30. doi:10.1016/j.jinsphys.2013.03.005

Vanden Broeck, J., 1996. G-Protein-Coupled Receptors in insect cells. International Review of Cytology 164, 189–268. https://doi.org/10.1016/S0074-7696(08)62387-6

Vanden Broeck, J., Veelaert, D., Bendena, W.G., Tobe, S.S., De Loof, A., 1996. Molecular cloning of the precursor cDNA for schistostatins, locust allatostatin-like peptides with myoinhibiting properties. Molecular and Cellular Endocrinology 122(2), 191-198. doi: 10.1016/0303-7207(96)03890-7.

Veelaert, D., Devreese, B., Vanden Broeck, J., Yu, C.G., Schoofs, L., Van Beeumen, J., Tobe, S.S., De Loof A., 1996a. Isolation and characterization of schistostatin-2(11-18) from the desert locust, *Schistocerca gregaria*: a truncated analog of schistostatin-2. Regulatory Peptides 67(3), 195-199. doi: 10.1016/s0167-0115(96)00131-0.)

Veelaert, D., Devreese, B., Schoofs, L., Van Beeumen, J., Vanden Broeck, J., Tobe, S.S., De Loof A., 1996b. Isolation and characterization of eight myoinhibiting peptides from the desert locust, *Schistocerca gregaria*: new members of the cockroach allatostatin family. Molecular and Cellular Endocrinology 122(2), 183-190. doi: 10.1016/0303-7207(96)03889-0.)

Veelaert, D., Schoofs, L., Verhaert, P., De Loof, A., 1997. Identification of two novel peptides from the central nervous system of the desert locust, *Schistocerca gregaria*. Biochemical and Biophysical Research Communications 241, 530–534. https://doi.org/10.1006/bbrc.1997.7843

Veelaert, D., Baggerman, G., Derua, R., Waelkens, E., Meeusen, T., Vande Water, G., De Loof, A., Schoofs, L.,

1999. Identification of a new tachykinin from the midgut of the desert locust, *Schistocerca gregaria*, by ESI-Qq-oa-TOF mass spectrometry. Biochemical and Biophysical Research Communications 266(1), 237-242. doi: 10.1006/bbrc.1999.1808

Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Frontiers in Physiology 5, 1–22. https://doi.org/10.3389/fphys.2014.00454

Veenstra, J.A., 2016. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. General and comparative endocrinology 230-231, 153-157. doi:10.1016/j.ygcen.2016.04.013

Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. PeerJ 7, e7144. https://doi.org/10.7717/peerj.7144

Veenstra, J.A., Šimo, L., 2020. The TRH-ortholog EFLamide in the migratory locust. Insect Biochemistry and Molecular Biology 116, 103281. https://doi.org/https://doi.org/10.1016/j.ibmb.2019.103281

Veenstra, J.A., Leyria, J., Orchard, I., Lange, A.B., 2021. Identification of Gonadulin and Insulin-Like Growth Factor from migratory locusts and their importance in reproduction in *Locusta migratoria*. Frontiers in Endocrinology 12, 693068.https://doi.org/10.3389/fendo.2021.693068

Veenstra (2021) Identification of cells expressing Calcitonins A and B, PDF and ACP in *Locusta migratoria* using cross-reacting antisera and in situ hybridization. bioRxiv https://doi.org/10.1101/2021.07.28.454216

Verdonck, R., De Haes, W., Cardoen, D., Menschaert, G., Huhn, T., Landuyt, B., Baggerman, G., Boonen, K., Wenseleers, T., Schoofs, L., 2016. Fast and reliable quantitative peptidomics with labelpepmatch. Journal of Proteome Research 15, 1080–1089. https://doi.org/10.1021/acs.jproteome.5b00845

Verlinden, H., Sterck, L., Li, J., Li, Z., Yssel, A., Gansemans, Y., Verdonck, R., Holtof, M., Song, H., Behmer, S.T., Sword, G.A., Matheson, T., Ott, S.R., Deforce, D., Van Nieuwerburgh, F., Van de Peer, Y., Vanden Broeck, J., 2020. First draft genome assembly of the desert locust, *Schistocerca gregaria*. F1000Research. https://doi.org/10.12688/f1000research.25148.1

Wang, Xianhui, Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Wang, Xuesong, Zhang, Y., Yang, M., Hao, S., Chen, B., Ma, Z., Yu, D., Xiong, Z., Zhu, Y., Fan, D., Han, L., Wang, B., Chen, Y., Wang, Junwen, Yang, L., Zhao, W., Feng, Y., Chen, G., Lian, J., Li, Q., Huang, Z., Yao, X., Lv, N., Zhang, G., Li, Y., Wang, Jian, Wang, Jun, Zhu, B., Kang, L., 2014. The locust genome provides insight into swarm formation and long-distance flight. Nature Communications 5, 2957. https://doi.org/10.1038/ncomms3957Wegener, C.,

Xie, J., Sang, M., Song, X., Zhang, S., Kim, D., Veenstra, J.A., Park, Y., Li, B., 2020. A new neuropeptide insect parathyroid hormone iPTH in the red flour beetle *Tribolium castaneum*. PLoS Genetics 16, e1008772. https://doi.org/10.1371/journal.pgen.1008772

Zeng, H., Qin, Y., Du, E., Wei, Q., Li, Y., Huang, D., Wang, G., Veenstra, J.A., Li, S., Li, N., 2021. Genomics and peptidomics-based discovery of conserved and novel neuropeptides in the American Cockroach. Journal of Proteome Research. https://doi.org/10.1021/acs.jproteome.0c00596

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. Molecular & Cellular Proteomics 11, M111.010587. https://doi.org/10.1074/mcp.M111.010587

Figures

Figure 1. MALDI-TOF MS^2 fragment ion spectrum of ACP from a preparation of the *corpus cardiacum* (CC, direct tissue profiling). Ion signals of *b*- and *y*-type fragment ions are labelled.

Figure 2. MALDI-TOF MS^1 spectra obtained by direct tissue profiling of various parts of the adult nervous system of *S gregaria*. A-B: CNS, C: stomatogastric nervous system; D-F: neurohemal organs. For the abbreviations of peptides, see Table 1. Labels in red are for ion signals mass-identical to peptides from neuropeptide precursors, labels in green are for ion signals mass-identical to peptides from neuropeptide-like precursors (function and receptors unknown). **A**) Mass spectrum of an antennal lobe (AL) preparation; ALs represent the primary olfactory brain neuropil. Note the abundance of peptides from neuropeptide-like precursors. *, TKRP-6; **, NVP; ***, kinin-3; ****, AstA-10. **B**) Mass spectrum of a preparation from the caudal neuropil region of the terminal ganglion (TermG) showing a peptidome similar to that of the AL preparation but additionally containing prominent ion signals indicating ACP and AstCC. The low ion signals of kinins and NPF1-PP are not labeled. **C**) Mass spectrum of a preparation from the frontal ganglion (FG) with an ion signal typical of hanSolin and particularly prominent ion signals of sNPF and MS. **D)** Mass spectrum from a preparation of a mainly glandular part of a *corpus cardiacum* (CC) as part of the RCC showing prominent ion signals of AKH1 and APRP1. Among the tissue analysed here, the CC spectra are the only ones showing enrichment of PK, PKL, IRP, and CRF-DH peptides. Note the absence or very low signal intensity of peptides from neuropeptide-like precursors, which has generally been observed in mass spectra of neurohemal organs. **E**) Mass spectrum from a preparation of a thoracic perisympathetic organ (tPSO1) containing mainly ion signals from peptides of FMRFa and NPF1 precursors. While FMRFa-2, 3, 4 have well-defined mature peptides, FMRFa-1 and 5 occur with N-termini of variable lengths. **F)** Mass spectrum from a preparation of an abdominal perisympathetic organ (aPSO3) containing ion signals from CAPA-peptides of both transcripts and different haplotypes.

Figure 3. MALDI-TOF MS^1 spectra (direct tissue profiling) from preparations of **A)** Lateral cardiac nerve showing ion signals mass-identical to peptides of both tryptoPK precursors. *, tPK2-7. **B)** Tracheal trunk of a 5th instar nymph suggesting the presence of ETH peptides. The identity of the peptides has been confirmed by subsequent MS^2 experiments using the same sample (not shown). For the abbreviations of peptides, see Table 1.

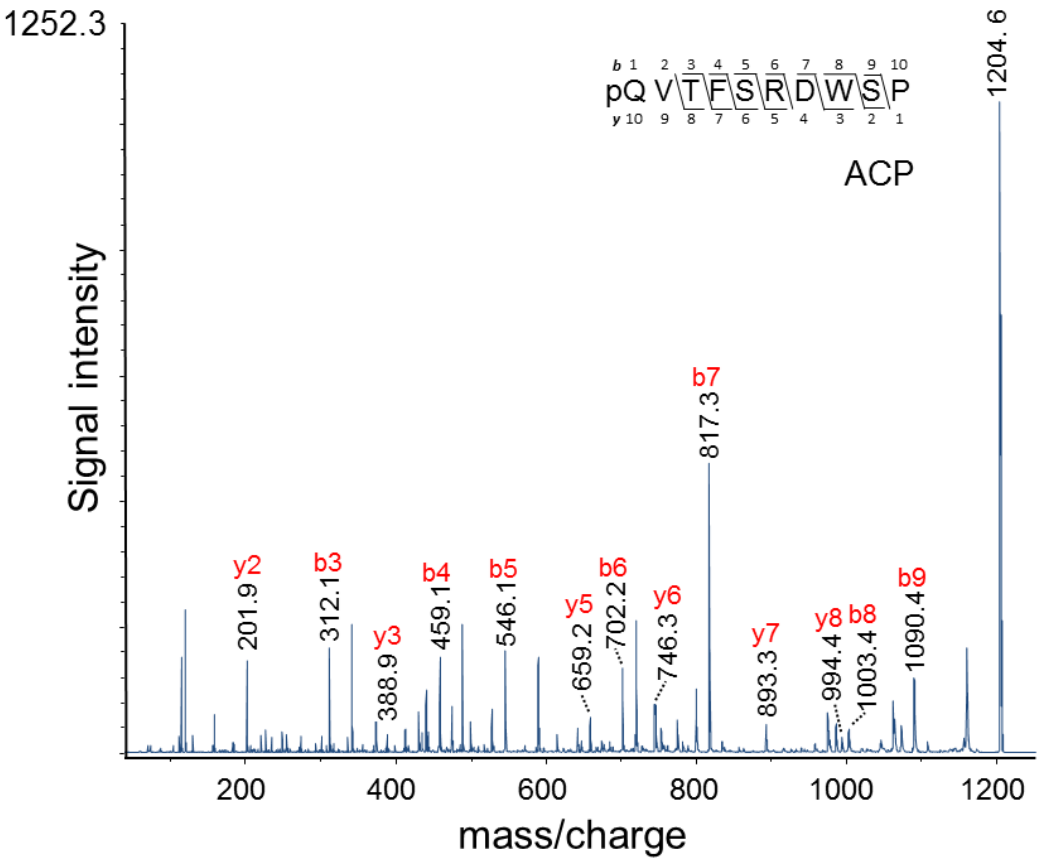

1252.3
Signal intensity
b 1 2 3 4 5 6 7 8 9 10
pQ V T F S R D W S P
y 10 9 8 7 6 5 4 3 2 1
ACP
1204. 6
b7
817.3
y2
201.9
b3
312.1
y3
388.9
b4
459.1
b5
546.1
y5
659.2
b6
702.2
y6
746.3
y7
893.3
y8
994.4
b8
1003.4
b9
1090.4
200
400
600
800
1000
1200
mass/charge

Figure 1

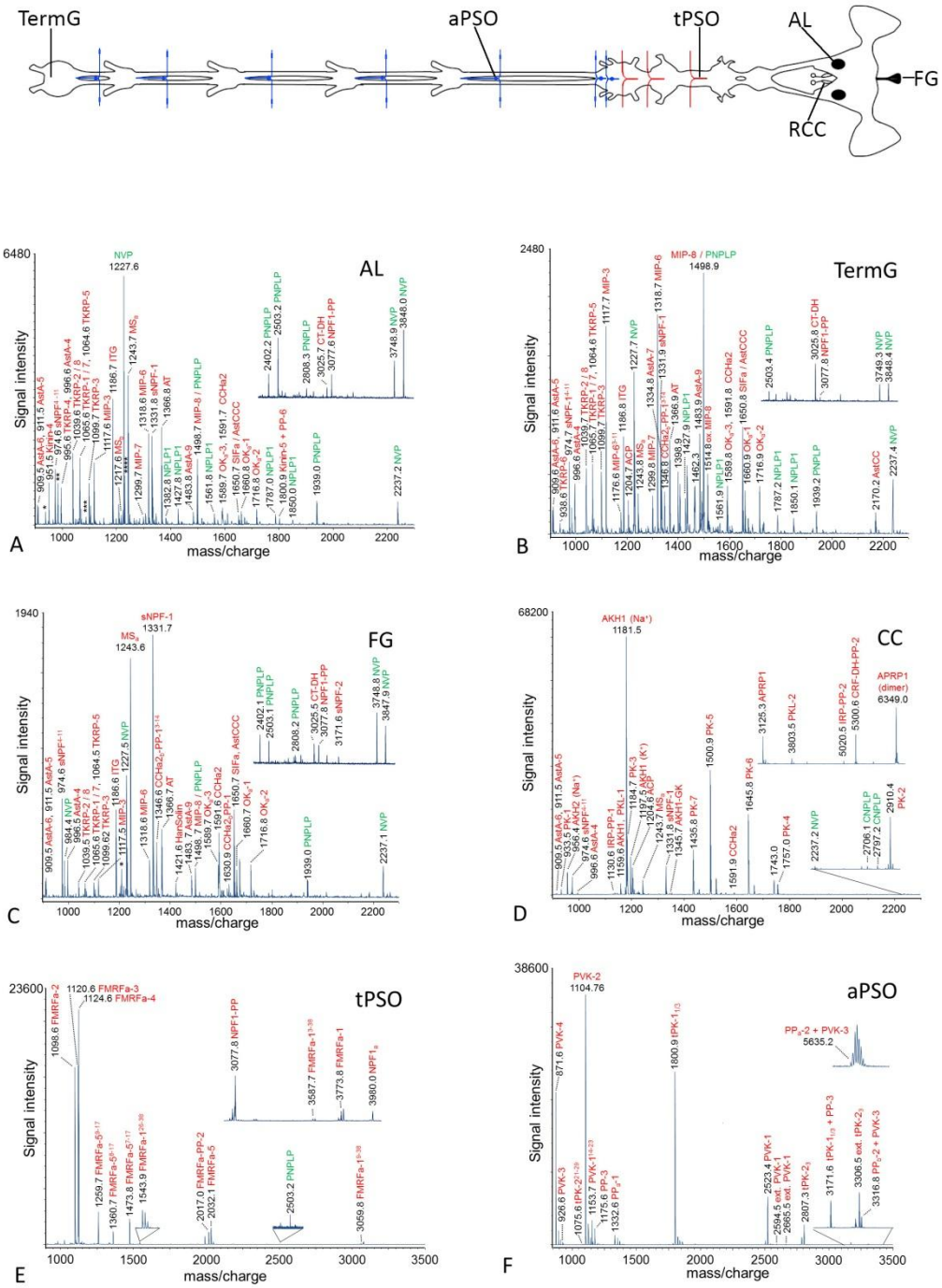

TermG
aPSO
tPSO
AL
FG
RCC
AL
TermG
FG
CC
tPSO
aPSO
Signal intensity
mass/charge
A
B
C
D
E
F

Figure 2

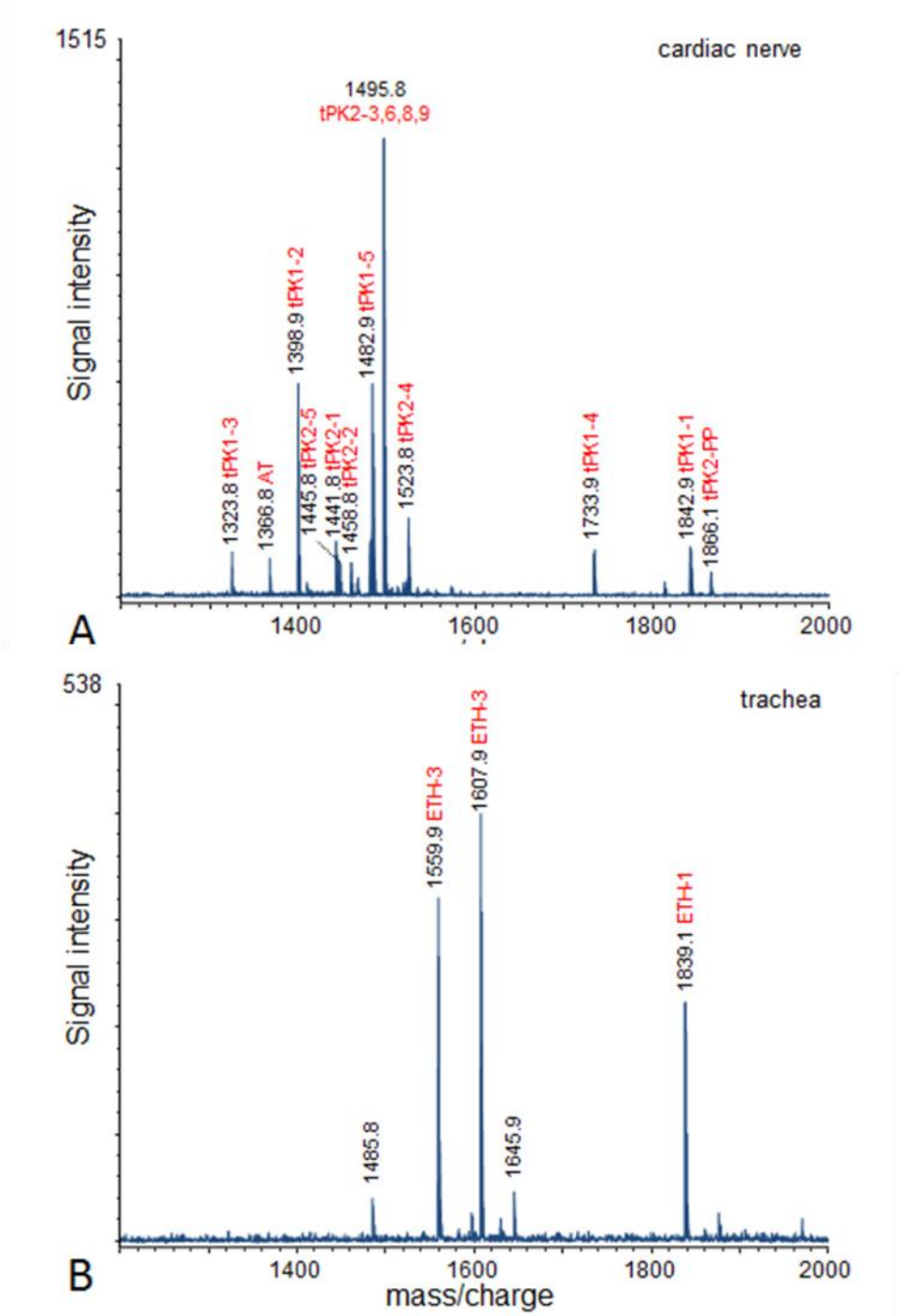

1515
cardiac nerve
1495.8
tPK2-3,6,8,9
Signal intensity
1398.9 tPK1-2
1482.9 tPK1-5
1323.8 tPK1-3
1366.8 AT
1445.8 tPK2-5
1441.8 tPK2-1
1458.8 tPK2-2
1523.8 tPK2-4
1733.9 tPK1-4
1842.9 tPK1-1
1866.1 tPK2-PP
A
1400
1600
1800
2000
538
trachea
1607.9 ETH-3
1559.9 ETH-3
1839.1 ETH-1
Signal intensity
1485.8
1645.9
B
1400
1600
1800
2000
mass/charge

Figure 3

Name	Abbreviation	Accession	AA	complete	NPs in precursor	MS^2
Adipokinetic hormone 1	AKH1	pending	63	+	1	+
Adipokinetic hormone 2	AKH2	pending	61	+	1	+
Adipokinetic hormone 3	AKH3	pending	73	+	1	-
Adipokinetic hormone / corazonin-related peptide	ACP	pending	88	+	1	+
Allatostatin A	AstA	pending	283	+	10	+
Allatostatin C (allele 1)	$AstC_1$	pending	81	+	1	-
Allatostatin C (allele 2)	$AstC_2$	pending	81	+	1	-
Allatostatin CC	AstCC	pending	142	+	1	+
Allatostatin CCC	AstCCC	pending	97	+	1	+
Allatotropin	AT	pending	117	+	1	+
Bursicon alpha	-	pending	> 148	-	1	-
Bursicon beta	-	pending	> 147	+	1	-
Calcitonin A	-	pending	157	+	1	-
Calcitonin B (allele 1)	-	pending	176	+	3	-
Calcitonin B (allele 2)	-	pending	176	+	3	-
Calcitonin-like diuretic hormone	CT-DH	pending	120	+	1	+
$CAPA_a$ (allele 1)	$CAPA_{a1}$	pending	276	+	7	+
$CAPA_a$ (allele 2)	$CAPA_{a2}$	pending	276	+	7	+
$CAPA_a$ (allele 3)	$CAPA_{a3}$	pending	276	+	6	+
$CAPA_b$	$CAPA_b$	pending	225	+	6	+
CCHamide 1	CCHa1	pending	71	+	1	+
CCHamide 2_a	$CCHa2_a$	pending	121	+	1	+
CCHamide 2_b	$CCHa2_b$	pending	137	+	1	+
CNMamide	CNMa	pending	167	+	1	-
Corazonin	Crz	pending	133	+	1	+
Corticotropin-releasing factor-like DH	CRF-DH	pending	185	+	1	+
Crustacean cardioactive peptide	CCAP	pending	149	+	1	+
Ecdysis-triggering hormone	ETH	pending	153	+	3	+
Eclosion hormone 1	EH1	pending	86	+	1	-
Eclosion hormone 2 (allele 1)	$EH2_1$	pending	83	+	1	-
Eclosion hormone 2 (allele 2)	$EH2_2$	pending	83	+	1	-
EFLamide	-	pending	> 78	-	≥1	-
Elevenin	-	pending	100	+	1	-
extended FMRFamide	FMRFa	pending	198	+	6	+
Glycoprotein hormone alpha 2	GPA2	pending	132	+	1	-
Glycoprotein hormone beta 5	GPB5	pending	153	+	1	-
Gonadulin	-	pending	139	+	1	-
HanSolin	-	pending	109	+	1	+
IDL-containing	IDL	pending	205	+	?	+
Inotocin	-	pending	169	+	1	-
Insect parathyroid hormone	iPTH	pending	160	+	1	-
Insulin-like growth factor	IGF	pending	223	+	1	-
Insulin-related peptide	IRP	pending	143	+	2	+
Ion transport $peptide_a$	ITP_a	pending	134	+	1	+
Ion transport $peptide_b$	ITP_b	pending	130	+	1	-

ITG-like (allele 1)	ITG_1	pending	222	+	1	+
ITG-like (allele 2)	ITG_2	pending	222	+	1	+
Kinin	-	pending	212	+	6	+
Myoinhibitory peptide$_a$	MIP_a	pending	230	+	8	+
Myoinhibitory peptide$_b$	MIP_b	pending	244	+	8	?
Myosuppressin$_a$	MS_a	pending	97	+	1	+
Myosuppressin$_b$	MS_b	pending	100	+	1	+
Natalisin	Nat	pending	332	+	9	+
Neuroparsin$_a$	-	pending	107	+	1	-
Neuroparsin$_b$	-	pending	103	+	1	-
Neuroparsin$_c$	-	pending	103	+	1	-
Neuroparsin$_d$	-	pending	104	+	1	-
Neuropeptide F1$_a$	$NPF1_a$	pending	98	+	1	+
Neuropeptide F1$_b$	$NPF1_b$	pending	148	+	1	+
Neuropeptide F2	NPF2	pending	99	+	1	+
Orcokinin$_a$	OK_a	pending	167	+	4	+
Orcokinin$_b$	OK_b	pending	340	+	20	-
Pigment dispersing factor	PDF	pending	89	+	1	+
Proctolin	-	pending	79	+	1	+
Prothoracicotropic hormone	PTTH	pending	177	+	1	+
Pyrokinin	PK	pending	> 150	-	7	+
Pyrokinin-like	PKL	pending	172	+	3	+
RFLamide	RFLa	pending	166	+	1	+
RYamide	RYa	pending	118	+	1	+
short Neuropeptide F$_a$	$sNPF_a$	pending	129	+	2	+
short Neuropeptide F$_b$	$sNPF_b$	pending	132	+	2	-
SIFamide	SIFa	pending	75	+	1	+
SMYamide	SMY	pending	73	+	1	-
Sulfakinin	SK	pending	110	+	2	+
Tachykinin-related peptide	TKRP	pending	299	+	9	+
Trissin	-	pending	114	+	1	-
Tryptopyrokinin 1	tryptoPK1	pending	269	+	5	+
Tryptopyrokinin 2	tryptoPK2	pending	295	+	10	+
Neuropeptide-like						
Agatoxin-like peptide$_a$	ALP_a	pending	107	+	?	(+)
Agatoxin-like peptide$_b$	ALP_b	pending	98	+	?	(+)
Agatoxin-like peptide$_c$	ALP_c	pending	116	+	?	(+)
Carausius neuropeptide-like precursor 1 (allele 1)	$CNPLP_1$	pending	404	+	?	+
Carausius neuropeptide-like precursor 1 (allele 2)	$CNPLP_2$	pending	404	+	?	-
Neuropeptide-like precursor 1	NPLP1	pending	570	+	?	+
NVP-like$_a$	NVP_a	pending	333	+	?	+
NVP-like$_b$	NVP_b	pending	262	+	?	-
Periplaneta neuropeptide-like precursor	PNPLP	pending	1272	+	?	+
Salivary gland salivation stimulating peptide	SGSSP	pending	> 340	-	?	-

Designation	peptide sequence	$[M+H]^+$, m/z	MALDI	Orbitrap	Original description *S. gregaria*
Adipokinetic hormone 1 (AKH1)					
AKH1	pQLNFTPNWGT-NH_2	1159.55	MS^2	MS^2	Stone et al. 1976
AKH1 (Na^+)	pQLNFTPNWGT-NH_2	1181.53	MS^2	-	
ext. AKH1	SAQLNFTPNWGT-NH_2	1334.65	MS^1	MS^2	
ext. AKH1	AQLNFTPNWGT-NH_2	1247.62	MS^1	MS^2	
AKH1-GKR	pQLNFTPNWGTGKR-OH	1501.75	MS^2	-	
AKH1-GK	pQLNFTPNWGTGK-OH	1345.65	MS^2	MS^2	
APRP1	Dimer AKH1-PPs	6248.89	MS^1	-	Hekimi et al. 1989
PP	DAADFGDPYSFLYRLIQAEARKMSGCSN-OH	3125.45	MS^2	MS^2	Hekimi et al. 1989
PP (sulf.)	DAADFGDPY(SO_3)SFLYRLIQAEARKMSGCSN-OH	3205.40	-	MS^2	
PP^{1-12}	DAADFGDPYSFL-OH	1317.56	MS^1	MS^2	Clynen & Schoofs 2009
AKH2					
AKH2 (Na^+)	pQLNFSTGW-NH_2	956.42	MS^2	-	Gäde et al. 1986
AKH2-GRR	pQLNFSTGWGRR-OH	1304.65	MS^1	-	
AKH2-GR	pQLNFSTGWGR-OH	1148.55	MS^1	MS^2	
PP (sulf.)	Y(SO_3)ADPNADPMAFLYKLIQIEARKLSGCSN-OH	3208.51	-	MS^2	
PP^{1-12}	YADPNADPMAFL-OH	1324.59	-	MS^2	Clynen & Schoofs 2009
Adipokinetic hormone/corazonin-related peptide (ACP)					
ACP	pQVTFSRDWSP-NH_2	1204.57	MS^2	-	Siegert 1999*
Agatoxin-like peptide (ALP)					
ALP	ACIRRGGTCDHRPKDCCYNSSCRCNLWGANCRCQRMGLFQKW-NH_2	4887.09	MS^1	-	Sturm et al. 2016
$PP\text{-}1^{1-24}$	GPYLDDPVPDDGVEDYSDGNLERL-OH	2650.18	-	MS^2	
Allatostatin A (AstA)					
ext. AstA-1	YKRLYDFGV-NH_2	1159.63	-	MS^2	(Vanden Broeck et al. 1996)
AstA-2	LPVYNFGL-NH_2	921.52	-	MS^2	Veelaert et al. 1996a*
ext. AstA-2	AYTYVSEYKRLPVYNFGL-NH_2	2182.13	MS^1	MS^2	Veelaert et al. 1996b*
AstA-3	ATGAASLYSFGL-NH_2	1156.60	-	MS^2	Veelaert et al. 1996b
AstA-4	GPRTYSFGL-NH_2	996.53	MS^1	MS^2	Veelaert et al. 1996b
AstA-5	GRLYSFGL-NH_2	911.51	MS^2	MS^2	Veelaert et al. 1996b
AstA-6	ARPYSFGL-NH_2	909.49	MS^1	MS^2	Veelaert et al. 1996b
AstA-7	AGPAPSRLYSFGL-NH_2	1334.72	MS^1	MS^2	Veelaert et al. 1996b
AstA-8	EGRMYSFGL-NH_2	1058.51	MS^1	MS^2	Veelaert et al. 1996b

AstA-9	PLYGGDRRFSFGL-NH_2	1483.78	MS^1	MS^2	(Vanden Broeck et al. 1996)
AstA-8 + 9	EGRMYSFGLGKRPLYGGDRRFSFGL-NH_2	2865.47	MS^1	MS^2	(Vanden Broeck et al. 1996)
AstA-10	APAEHRFSFGL-NH_2	1230.64	MS^1	MS^2	Veelaert et al. 1996b
Allatostatin CC (AstCC)					
AstCC	GQKTGQYWRCYFNAVTCF-OH	2169.96	MS^1	-	
PP	TALLLDRLMVDLKHLMDKDRGEAQNPIDSGSSIGRMALQ-OH	4308.23	-	MS^2	
Allatostatin CCC (AstCCC)					
AstCCC	SYWKQCAFNAVSCF-NH_2	1650.71	MS^2	MS^2	
PP	EPLGQQPSDKARLLNELDLVDDDGSIETALINYLFAKQVVNRLRAQMDVSDLQ-OH	5954.06	MS^1	MS^2	
ext. PP	RAEPLGQQPSDKARLLNELDLVDDDGSIETALINYLFAKQVVNRLRAQMDVSDLQ-OH	6181.20	-	MS^2	
Allatotropin (AT)					
AT	GFKNVALSTARGF-NH_2	1366.76	MS^1	MS^2	Paemen et al. 1991*
PP1	APAAHYGRGSRPRTI-OH	1609.87	MS^1	MS^2	
ext. PP-1	AAPAAHYGRGSRPRTI-OH	1680.90	MS^1	MS^2	
PP-1^{2-15}	PAAHYGRGSRPRTI-OH	1538.83	MS^1	MS^2	
PP-1 + AT	APAAHYGRGSRPRTIRGFKNVALSTARGF-NH_2	3113.71	-	MS^2	
PP-2	DGNQLEAALADRDTTLPDSFPVEWFAAEMQNNPELARMIVSKFVDANQDGELTAEELLRPTY-OH	6952.34	-	MS^2	
Calcitonin-like diuretic hormone (CT-DH)					
CT-DH	GLDLGINRGFSGAQAAKHLMGLAAAQYAAGP-NH_2	3025.58	MS^1	MS^2	
PP-2	DTHQQPQQPAGHTSA-OH	1602.73	-	MS^2	
CAPA (periviscerokinins/PVKs, pyrokinin/PK, tryptopyrokinin/tryptoPK)					
PVK-1	AEDGDKGISKLKKTSSLFPHPRI-NH_2	2523.40	MS^2	MS^2	
ext. PVK-1	AAEDGDKGISKLKKTSSLFPHPRI-NH_2	2594.44	MS^2	MS^2	
ext. PVK-1	AAAEDGDKGISKLKKTSSLFPHPRI-NH_2	2665.48	MS^1	MS^2	
PVK-1^{2-23}	EDGDKGISKLKKTSSLFPHPRI-NH_2	2452.37	MS^1	MS^2	
PVK-1^{13-23}	KTSSLFPHPRI-NH_2	1281.74	MS^2	MS^2	
PVK-1^{14-23}	TSSLFPHPRI-NH_2	1153.65	MS^2	MS^2	Veelaert et al. 1997*
PVK-2_a	AAGLFQFPRV-NH_2	1104.63	MS^2	-	Predel & Gäde 2002
PVK-2_a^{3-10}	GLFQFPRV-NH_2	962.56	MS^1	MS^2	
PVK-3	KGLVANARV-NH_2	926.59	MS^1	MS^2	

PVK-3$^{2-9}$	GLVANARV-NH_2	798.49	MS^1	MS^2	
PP_a-2 + PVK-3	AFIDAPLPFPLSVVGPLRLAPSDKDADAEAATVAEQPSPFEGHKRKGLVANARV-NH_2	5635.02	MS^2	MS^2	
PP_b-2 + PVK-3	SEFINHEAVAEQPSPFEGHKRKGLVANARV-NH_2	3316.73	MS^1	MS^2	
CAPA-tryptoPK-$1_{1/3}$	DGAETPGAAASLWFGPRV-NH_2	1800.90	MS^1	MS^2	Clynen et al. 2003a*
CAPA-tryptoPK-1_2	DGADTPGAAASLWFGPRV-NH_2	1786.89	MS^1	MS^2	
CAPA-tryptoPK-$1_{1/3}$ + PP-3	DGAETPGAAASLWFGPRVGRAGLGQDETRAGT-OH	3171.56	MS^1	MS^2	
CAPA-tryptoPK-1_2 + PP-3	DGADTPGAAASLWFGPRVGRAGLGQDETRAGT-OH	3157.54	MS^1	MS^2	
PVK-4	GLLAFPRV-NH_2	871.55	MS^1	MS^2	Clynen et al. 2003a
ext. PVK-4	RGLLAFPRV-NH_2	1027.65	MS^2	MS^2	
PP-3 + PVK-4	AGLGQDETRAGTKRRGLLAFPRV-NH_2	2468.40	MS^1	MS^2	
CAPA-tryptoPK-2	GHAGSSSSSSSGDGDGARDSLWFGPRV-NH_2	2635.19	MS^1	MS^2	
ext. CAPA-tryptoPK-2	GHAGSSSSSSSGDGDGARDSLWFGPRVGRRE-OH	3134.44	MS^1	MS^2	
CAPA-tryptoPK-2_3	GHAGSSSSSSSGDGDGDGARDSLWFGPRV-NH_2	2807.24	MS^2	MS^2	
ext. CAPA-tryptoPK-2_3	GHAGSSSSSSSGDGDGDGARDSLWFGPRVGRRE-OH	3306.49	MS^1	MS^2	
CAPA-tryptoPK-2$^{21-29}$	DSLWFGPRV-NH_2	1075.57	MS^1	-	
CAPA-PK	SLRLRLPAAAWLAAGDVGNGKGDFTPRL-NH_2	2921.62	MS^1	MS^2	
PP_a-1	SEFINHEAGSGGQ-OH	1332.58	MS^1	MS^2	
PP_b-1	SEFINHEAVAEQPSPFEGH-OH	2124.96	MS^1	MS^2	
PP_a-2	AFIDAPLPFPLSVVGPLRLAPSDKDADAEAATVAEQPSPFEGH-OH	4443.25	MS^2	MS^2	
PP-3	AGLGQDETRAGT-OH	1175.56	MS^1	MS^2	
***Carausius* neuropeptide-like precursor**					
	SRSVESVERPLMVSVSESSGKNGRY-OH	2740.36	MS^1	MS^2	
	EVGALPSMPSVTKAMEYS-OH	1896.91	MS^1	MS^2	
	RPEMGSSGFHGNMFSNGFGEFWPM-OH	2706.13	MS^2	MS^2	
	SNTLDMDGDVGVDKLLPM-OH	1919.90	MS^1	MS^2	
	RPEMDSLGFHGDTFHNGFGDFWPM-OH	2797.19	MS^1	MS^2	
	STPEENFATGRSDPKTEQNYTHSKECCGDNEKN-OH	3714.56	-	MS^2	
CCHamide 1 (CCHa1)					
CCHa1	SCLSYGHSCWGAH-NH_2	1404.56	MS^1	MS^2	
PP$^{1-24}$	SGGGGASPRLGLVPDDARWTVSRL-OH	2424.27	MS^1	MS^2	
CCHamide 2 (CCHa2)					
$CCHa2_a$	KRGCMAFGHSCFGGH-NH_2	1591.68	MS^1	-	
$CCHa2_a^{3-15}$	GCMAFGHSCFGGH-NH_2	1307.48	-	MS^2	
$CCHa2_b$	RAGGCMAFGHSCFGGH-NH_2	1591.64	MS^1	MS^2	

CCHa2$_b$ + PP$_b$-1	KRSYGVRRPGDIQVRRAGGCMAFGHSCFGGH-NH$_2$	3359.63	-	MS^2	
PP$_b$-1	KRSYGVRRPGDIQV-OH	1630.91	MS^1	MS^2	
PP$_b$-1$^{3-14}$	SYGVRRPGDIQV-OH	1346.72	MS^1	MS^2	
CCHa2-PP	ADMEPAAEGVEGAEEAAAVAEAEAAAAAALLDDAASPQRFRLSPFLRQWLQRAY QQQSADSQTVEVK-OH	7100.47	-	MS^2	
Corazonin (Crz)					
Crz	pQTFQYSHGWTN-NH$_2$	1350.58	MS^2	MS^2	Tawfik et al. 1999
PP	AGSSSAPGALLPPGRLPPPAAASDMDAQPCRVRCLRLLLQGGAVPQLYVPPELW QQVDEEGDNMAARQRGGGARLRHALPPPGAAAAVDSDEDM-OH	9697.84	-	MS^2	
CRF-like diuretic hormone (CRF-DH)					
CRF-DH	MGMGPSLSIVNPMDVLRQRLLLEIARRRLRDAEEQIKANKDFLQQI-NH$_2$	5360.93	MS^1	MS^2	(Van Wielendaele et al. 2012)
CRF-DH$^{28-46}$	RLRDAEEQIKANKDFLQQI-NH$_2$	2314.27	MS^1	MS^2	
CRF-DH$^{31-46}$	DAEEQIKANKDFLQQI-NH$_2$	1888.98	MS^1	MS^2	
OMP-1	YYEAPPDGQRLLLQAAPAAAPAAASWPHQQRRQAIDEFAAAAAAAADAQYQDE EEDGARRV-OH	6534.16	MS^1	MS^2	
OMP-1$^{1-30}$	YYEAPPDGQRLLLQAAPAAAPAAASWPHQQ-OH	3188.59	MS^1	MS^2	
OMP-1$^{33-62}$	pQAIDEFAAAAAAAADAQYQDEEEDGARRV-OH	3035.35	MS^2	MS^2	
PP	SPHAGGAANDAGADAPPFGLRAAAERSASDISKDWASSDSRWNNQFTVRQS-OH	5300.49	MS^1	MS^2	
Crustacean cardioactive peptide (CCAP)					
CCAP-PP	AAAPPTPVQGMKPW-OH	1449.76	-	MS^2	
Ecdysis-triggering hormone (ETH)					
ETH-1	DEGANLFLKASRSVPHV-NH$_2$	1838.99	MS^1	MS^2	
ETH-2	SDFFLKTAKSVPRI-NH$_2$	1607.93	MS^2	MS^2	(Lenaerts et al. 2017)
ETH-3	SDLFLKSAKSVPRI-NH$_2$	1559.93	MS^2	MS^2	(Lenaerts et al. 2017)
ext. ETH-3	RSDLFLKSAKSVPRI-NH$_2$	1716.03	-	MS^2	
PP-1	TNLAAIEAQDGSEWLWPGGADAMPAPV-OH	2767.30	-	MS^2	
PP-2	pQAYYVRKDGQPVMWSDVARDVEENPDLWPWNDFDAGNTREVDNS-OH	5167.32	-	MS^2	
Extended FMRFamide (FMRFa)					
FMRFa-1	DADAEAAAVDDDGAGGEGDGDGELGLVQTTPRSNFLRL-NH$_2$	3773.75	MS^1	-	
FMRFa-1$^{3-38}$	DAEAAAVDDDGAGGEGDGDGELGLVQTTPRSNFLRL-NH$_2$	3587.68	MS^1	-	
FMRFa-1$^{9-38}$	VDDDGAGGEGDGDGELGLVQTTPRSNFLRL-NH$_2$	3059.47	MS^2	MS^2	
FMRFa-1$^{26-38}$	LVQTTPRSNFLRL-NH$_2$	1543.91	MS^1	MS^2	
FMRFa-2	AGGAHSAFLRL-NH$_2$	1098.62	MS^2	MS^2	

FMRFa-3	DRASSGFLRL-NH_2	1120.62	MS^2	MS^2	
FMRFa-4	GSERNFLRF-NH_2	1124. 60	MS^2	MS^2	
FMRFa-5	SGMADPLTRHDRNFIRF-NH_2	2032.03	MS^2	MS^2	
FMRFa-5^{7-17}	LTRHDRNFIRF-NH_2	1473.82	MS^2	MS^2	
FMRFa-5^{8-17}	TRHDRNFIRF-NH_2	1360.73	MS^2	MS^2	
FMRFa-5^{9-17}	RHDRNFIRF-NH_2	1259.69	MS^2	MS^2	
FMRFa-5^{10-17}	HDRNFIRF-NH_2	1103.59	MS^1	MS^2	
FMRFa-6	SGRSGTAQHQAGAEPALWAALLDPRLAVLPAPAGADDADVDGAAGSGRGRVST RSFLRL-NH_2	5879.06	MS^1	MS^2	
PP	SRQDGEQVAEDDELLSRA-NH_2	2016.96	MS^2	MS^2	
HanSolin					
hanSolin	GSRRPLGQPLRW-NH_2	1421.82	MS^1	MS^2	
PP	PASQEALSAET-OH	1103.52	-	MS^2	
IDL containing (IDL)					
IDL	IDLSRLYGHLSA-OH	1344.72	MS^1	MS^2	
IDL^{2-12}	IDLSRLYGHLS-OH	1273.69	-	MS^2	
Insulin-related peptide (IRP)					
IRP (B-chain)	SGAPQPVARYCGEKLSNALKIVCRGNYNTMF-OH	3385.66	MS^1	MS^2	Badisco et al. 2008
PP-1	pQSDLFLLSPK-OH	1130.61	MS^2	MS^2	Clynen et al. 2003c
PP-2	ASQDVSDAESEDNYWSQSADEEVEAPALPPYPVLARPSAGGLLTAAVF-OH	5020.36	MS^1	MS^2	
Ion transport peptide (ITP)					
ITP-PP	SPLDPHHLA-OH	986.51	-	MS^2	(Meredith et al. 1996)
ITP_a^{62-79}	MDKIQSWIKQIHGAEPGV-OH	2037.06	-	MS^2	
ITG-like (ITG)					
ITG^{1-48}	WGGLFNRFSPEMLSNLGYGGHGYGAYRSSQPLLQRFHNPVEVFQELQE-OH	5512.67	-	MS^2	
PP	ITGKVASFNHI-OH	1186.66	MS^2	MS^2	
Kinin					
kinin-1	SFSAWG-NH_2	653.30	MS^2	-	
kinin-2, 5	AFSSWG-NH_2	653.30	MS^2	-	Schoofs et al. 1992
kinin-3	SKQSFYSWG-NH_2	1088.52	MS^2	MS^2	
kinin-4	GGVRFSSWG-NH_2	951.48	MS^1	-	
kinin-5 + PP-6	AFSSWGGKRFGDGASAAE-OH	1800.83	MS^1	MS^2	
kinin-6	NGNGPPFFPWG-NH_2	1188.55	MS^2	MS^2	
PP-1	AERCCGDAEPWQGQARS-OH	1861.76	MS^1	MS^2	

PP-1$^{2-17}$	ERCCGDAEPWQGQARS-OH	1792.75	-	MS^{2}	
PP-3	AAAGWQAED-OH	918.40	MS^{1}	MS^{2}	
PP-4	GVGGDAEADGQAPEDAVE-OH	1686.71	-	MS^{2}	
PP-8	LDLYNYRPVQL-OH	1393.75	MS^{1}	-	
PP-7 + 8	AFSNWGESLANRRLDLYNYRPVQL-OH	2882.47	MS^{1}	MS^{2}	
PP-7 + 8 + kinin-6	AFSNWGESLANRRLDLYNYRPVQLRNGNGPPFFPWG-NH_2	4208.11	-	MS^{2}	
Myoinhibitory peptide (MIP)					
MIP-1, MIP_a-2	AWQDLNAGW-NH_2	1059.50	MS^{1}	MS^{2}	Schoofs et al. 1991
MIP-3	GWRDLQSAW-NH_2	1117.55	MS^{2}	MS^{2}	
MIP-4	AWSNLHGAW-NH_2	1040.51	MS^{2}	MS^{2}	
MIP-5	AWSSLHNTW-NH_2	1100.53	MS^{2}	MS^{2}	
MIP-6	AADWRSFHGSW-NH_2	1318.61	MS^{2}	MS^{2}	
MIP-6$^{3-11}$	DWRSFHGSW-NH_2	1176.53	MS^{1}	MS^{2}	
MIP-7 (E)	EPGWTNLKGLW-NH_2	1299.68	MS^{2}	MS^{2}	
MIP-7 (pQ)	pQPGWTNLKGLW-NH_2	1282.65	-	MS^{2}	
MIP-8	AGPSNWNRLPAMW-NH_2	1498.73	MS^{1}	MS^{2}	
PP-1	APQAPAADTVSAAAPAAAAASPAASDEDKSDEE-OH	3053.38	MS^{1}	MS^{2}	
PP-1$^{3-33}$	pQAPAADTVSAAAPAAAAASPAASDEDKSDEE-OH	2868.26	-	MS^{2}	
PP-2	GGDDAATWPELPDQTVAEQEGDQESAQLVPLPLQVPLPLQLQLQLQGDEGEPG DEE-OH	5989.86	-	MS^{2}	
Myosuppressin (MS)					
MS_a	PDVDHVFLRF-NH_2	1243.66	MS^{2}	MS^{2}	Robb et al. 1989
MS_b (E)	EDVGHVFLRF-NH_2	1217.64	MS^{2}	MS^{2}	
MS_b (pQ)	pEDVGHVFLRF-NH_2	1200.61	-	MS^{2}	
PP_a	NAVPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAVRDPELFVED-OH	6022.98	-	MS^{2}	
PP_a^{3-56}	VPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAVRDPELFVED-OH	6208.06	-	MS^{2}	
PP_b	NAVPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAEAHRQSVRPL PAGI-OH	6520.28	-	MS^{2}	
PP_b^{3-60}	VPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAEAHRQSVRPLPA GI-OH	6335.21	-	MS^{2}	
Natalisin (Nat)					
Nat-3	REMSASSASAEGPFWAAR-NH_2	1909.90	MS^{1}	MS^{2}	
Nat-4	EMEFFWPAR-NH_2	1211.57	MS^{1}	-	
Nat-5	AGLRSYLASMSGGQSFVPAR-NH_2	2054.06	MS^{1}	MS^{2}	

Nat-6	ASLGSESVEGPFWAPR-NH_2	1688.84	MS^1	MS^2	
Nat-6 + PP-5	ETGEVELASRRASLGSESVEGPFWAPR-NH_2	2916.46	MS^1	MS^2	
Nat-7	GQAARLLRLRLASAPAGSAFWPAR-NH_2	2535.45	MS^1	MS^2	
Nat-7^{11-24}	LASAPAGSAFWPAR-NH_2	1400.74	MS^1	MS^2	
Nat-8 + 9	AGPSSASSSAEGPFWATRGRSLAAQHSFWPAR-NH_2	3329.63	MS^1	MS^2	
Nat-9	SLAAQHSFWPAR-NH_2	1369.71	MS^1	MS^2	
PP-1	AGVPEAGLADDGRS-OH	1314.63	-	MS^2	
PP-4	GSPHAGVS-OH	711.34	-	MS^2	
Neuropeptide F1 (NPF1)					
$NPF1_a$ (pQ)	pQQAAADGNKLEGLADALKYL*QEL*DRYYSQVARPRF-NH_2	3980.04	MS^1	MS^2 (pt.)	Van Wielendaele et al. 2013
$NPF1_a^{27-35}$	YSQVARPRF-NH_2	1122.61	MS^1	MS^2	Schoofs et al. 2001
$NPF1_b^{37-85}$	SGGGASLSVRSPLDTLNIAEHLRGVEKMVRMLQLQEYDRMYTPRNRPRF-NH_2	5643.93	-	MS^2	
PP	SELRPDVVDDVIPEETSADKFWRRFA-OH	3077.53	MS^1	-	
PP^{1-22}	SELRPDVVDDVIPEETSADKFW-OH	2547.22	MS^1	MS^2	
Neuropeptide F2 (NPF2)					
$NPF2^{1-31}$	LPAGADAGQQRPERPPMFTSPEELRNYLTQL-OH	3482.75	-	MS^2	
$NPF2^{35-43}$	YASLGRPRF-NH_2	1065.59	-	MS^2	
PP^{1-25}	GSAAAAAFRAASRLPVPPPDAYEQL-OH	2526.31	-	MS^2	
Neuropeptide-like precursor 1					
	LRPPQAVAEPGDGLMVP-NH_2	1745.94	-	MS^2	
	EPGDGLMVP-NH_2	913.44	-	MS^2	
	YVAALARNGELPLYGGWRKKQQRGDKRYTNRY-NH_2	3827.05	-	MS^2	
	YVAALARNGELPLYGGWRKKQQRGD-OH	2846.52	MS^1	MS^2	
	YVAALARNGELPLYGGWRKKQQ-OH	2518.37	MS^1	MS^2	
	YVAALARNGELPLYGGW-OH	1849.95	MS^2	MS^2	
	QQRGDKRYTNRY-NH_2	1583.81	-	MS^2	
	pQQRGDKRYTNRY-NH_2	1566.78	-	MS^2	
	YTNRY-NH_2	715.35	MS^1	-	
	SVASLARAGALLPGT-NH_2	1382.81	MS^1	MS^2	
	NIAAMAKNGLLSPSGPVLLDGDGSGNGGE-OH	2711.33	-	MS^2	
	SVGALARGGLLPTP-NH_2	1307.78	MS^1	MS^2	
	HIGSLARDYSLPSF-NH_2	1561.81	MS^1	MS^2	

	NLGSLARSGGLSNVRYVTTKKDDAERVGGHSEAD-OH	3559.78	-	MS^2	
	NLGSLARSGGLSNVRYVTT-OH	1965.05	MS^1	MS^2	
	DDAERVGGHSEAD-OH	1357.56	MS^1	MS^2	
	NIASFMRSRGSSLVE-OH	1653.84	MS^1	MS^2	
	YLASLVRSHGLPYPLTKKEDDGPGEI-OH	2855.49	-	MS^2	
	YLASLVRSHGLPYPLT-OH	1786.98	MS^1	MS^2	
	EDDGPGEI-OH	831.34	-	MS^2	
	NVGALARNWMLPS-NH_2	1427.75	MS^1	MS^2	
	ASDDDQEVDKRYLASVLRQ-NH_2	2207.10	-	MS^2	
	ASDDDQEVD-OH	993.36	-	MS^2	
	YLASVLRQ-NH_2	948.57	MS^1	MS^2	
	SDGFRQNSDGAQQADHEEE-OH	2119.85	MS^1	MS^2	
	HLGSLAKSGMAIH-OH	1321.70	-	MS^2	
	TSRSAGSDGQAFLQQQQQQEQGGAHAQDAAGS-OH	3244.46	-	MS^2	
	HAYLLPPAPPQSLAPAPGEFPMPVLQNNDDALDYGDLL-OH	4057.01	-	MS^2	
	FLGVPPAAADY-NH_2	1119.58	-	MS^2	
	HIGALARLGWLPSFRAASARSGRSAGSRS-NH_2	2994.64	-	MS^2	
	HIGALARLGWLPSFRAASARS-NH_2	2236.26	MS^1	MS^2	
	ATRSHSADGPWPAELQQA-OH	1921.91	MS^1	MS^2	
NVP-like					
	YGDPSAVNQYRYYGGANERRPDGAEGAFAPPS-OH	3432.56	-	MS^2	
	SSSFRPMVPHALELSGVGPRL-OH	2237.19	MS^2	MS^2	
	DSRWNGYSKD-OH	1227.54	MS^2	MS^2	
	DSRWNGYS-OH	984.42	MS^1	MS^2	
	DVTQPARGDIHYLAQLLGPSHRDQQIPLFHRVAV-OH	3848.05	MS^1	MS^2	
	DVTQPARGDIHYLAQLLGPSHRDQQIPLFHRVA-OH	3748.98	MS^1	MS^2	
Orcokinin$_a$ (OK$_a$)					
Orcomyotropin-like	SGLDSLSGATFGEQ-OH	1368.63	-	MS^2	
OK_a-1	NFDEIDRSGFNFSI-OH	1660.76	MS^2	MS^2	
OK_a-2	NFDEIDRSGFDRFV-OH	1716.80	MS^2	MS^2	
OK_a-1 + 2	NFDEIDRSGFNSFIKKNFDEIDRSGFDRFV-OH	3614.73	-	MS^2	
OK_a-3	NFDEIDRSGFSGFV-OH	1589.72	MS^1	MS^2	Hofer et al. 2005
PP_a-1	VPAPQMVSSGFQQYRDEPNDVEEGLVRHLDNIGGGHLLRNLDGLGGGHLLRQT-OH	5761.90	-	MS^2	
PP_a-1$^{28-38}$	HLDNIGGGHLL-OH	1145.61	-	MS^2	

PP_a-1$^{40-53}$	NLDGLGGGHLLRQT-OH	1450.78	MS^1	MS^2	
PP_a-2	LDSLSGITFGSQ-OH	1224.62	MS^1	MS^2	
PP_a-3	NAPMLARHYDQGDH-OH	1624.73	MS^1	MS^2	
***Periplaneta* neuropeptide-like precursor**					
	pQARGDSLQAALDAVT-OH	1498.75	MS^1	MS^2	
	pQARGDSLQAALDAVTRRQ-OH	1939.015	MS^2	MS^2	
	DLTLPARPSGYYALSQYRHQAAPATDDDIAF-OH	3423.66	-	MS^2	
	LDTGRDFTGYGQPENIGAGYQKTISSGAAPYPAVLSDVPVHPPPNQLENML-OH	5382.64	-	MS^2	
	YMKDAYVKDGDEEEGNYYQYGMDTDA-OH	3070.22	-	MS^2	
	LFRERENDDSGQDVEYHQVLPTYYDSENKQENVLLPSL-OH	4627.19	-	MS^2	
	FRERFLTRPEIDQVVELENMRRYAAKAIAKQLETDEEEKLEN-OH	5107.62	MS^1	MS^2	
	NNDEEEEYLSLLRNLWEKYKEAKPQLVDFDDLTQNDIQEILSSLRNDGSSLH-OH	6164.00	-	MS^2	
	QYGYGSGFDIFNNAGLMGQWGTGANNFA-OH	2957.29	-	MS^2	
	pQYGYGSGFDIFNNAGLMGQWGTGANNFA-OH	2940.27	-	MS^2	
	NKQRVEGPGQQGANFLYSL-OH	2106.07	-	MS^2	
	KFVAPEVNREAVETLKDNEGIELPDERDEDVLRLASGFARNNPEEL	5223.62	-	MS^2	
	MQIYGRPATMDE-OH	1411.63	-	MS^2	
	IYSPNQETYQTLSLETPDIGSRATSTKHFSSLARDVNEYQQLSPPNY-OH	5346.58	-	MS^2	
	MVIPERTSNKRFIYEAKRKRYPVT-OH	2982.64	-	MS^2	
	MVIPERTSNKRFIYEA-OH	1954.02	-	MS^2	
	MVIPERTSN-OH	1046.53	-	MS^2	
	SSNFYASPPMLHHKSFNSEGIKDTN-OH	2808.30	MS^1	MS^2	
	SSPITGVTDPKVAQELNQIFSSSVTHDDSPKASE-OH	3571.74	-	MS^2	
	TAPETVETKDGDGNMIDDNWLLNQYYRTLAMVSNPL-OH	4084.93	MS^1	MS^2	
	MVSHSHANSPNNNDN-OH	1637.67	MS^1	MS^2	
	DTGKQQAYNTDIFSRSSQREAT-OH	2503.18	MS^1	MS^2	
	SLNTKEDTSIDDMDTKLRNMEDL-OH	2684.24	-	MS^2	
	IVNEAVKYTGSHEGTQDPKEIQEMKDKIMSRLAAAYSLEKMRLAL-OH	5092.63	-	MS^2	
	KSSLQAQMMSKYNPANLKSSLSNDNTQEESKM-OH	3589.69	-	MS^2	
	KSSLQAQMMSKYNPANLKSSLSNDNTQEES-OH	3330.56	-	MS^2	
	DTGKQQAYNTDIFSRSSQREA-OH	2402.13	MS^1	MS^2	
	VAVKKEKAEDDKHNDDGD-OH	2012.95	-	MS^2	
Prothoracicotropic hormone (PTTH)					
PTTH-PP	TSLEDRLRPLWTEAEAEAAAAALAGAAGAAGPVGACSDPEDCAFHRS-NH_2	4679.20	-	MS^2	
Pigment dispersing factor (PDF)					

PDF	NSEIINSLLGLPKLLNDA-NH_2	1923.09	-	MS^2	
PP^{1-19}	TQYEEEKYQENEVRYGREL-OH	2463.14	-	MS^2	
Proctolin					
Proctolin	RYLPT-OH	649.37	MS^1	-	
PP^{1-19}	SSASLEDRLDRLRDLINDL-OH	2201.15	-	MS^2	
PP^{20-30}	VESERPSARMAPP-OH	1426.71	-	MS^2	
Pyrokinin (PK)					
PK-1	EGDFTPRL-NH_2	933.48	MS^2	MS^2	Clynen et al. 2003a
PK-2	ESAEQGAAPQWQSAEEQVLSGPFVPRL-NH_2	2910.44	MS^2	MS^2	
PK-3	GAAPAAQFSPRL-NH_2	1184.65	MS^2	MS^2	Veelaert et al. 1997*
PK-4	DPPADGLVWLPLVPRL-NH_2	1757.01	MS^2	MS^2	Clynen et al. 2003b*
PK-5	RPLPAPAAPFVPRL-NH_2	1500.92	MS^2	MS^2	Clynen et al. 2003b
PK-6	DSSEDWAQPFVPRL-NH_2	1645.80	MS^2	MS^2	Clynen et al. 2003b
PK-7	LQQYGMPFSPRL-NH_2	1435.75	MS^2	MS^2	Clynen et al. 2003b
$PK\text{-}7^{2-12}$	pQQYGMPFSPRL-NH_2	1305.64	MS^1	MS^2	
$PK\text{-}7^{3-12}$	pQYGMPFSPRL-NH_2	1177.58	MS^1	MS^2	(Clynen et al. 2003b)
PP	DAPDQLQADEQ-OH	1229.52	-	MS^2	
Pyrokinin-like (PKL)					
PKL-1	pQSMPTFTPRL-NH_2	1159.59	MS^2	MS^2	(Clynen et al. 2003b)*
PKL-2	DSAGDDLAEEEAEGDGDADGGHGQGQLQLAPPFWPRP-NH_2	3803.68	MS^2	MS^2	
PKL-3	HAPPLPLTPRL-NH_2	1210.74	MS^2	MS^2	
PP-2	DALGDMMLAHVIEHPWIVMPLAAVGSAPVGAAPPPAS-OH	3688.87	-	MS^2	
PP-3	AAHQATPAASGEPTSA-OH	1466.69	-	MS^2	
RFLamide (RFLa)					
$RFLa\text{-}PP\text{-}1^{1-25}$	APAPAPQAAARSPTEITADPEQLEL-OH	2544.29	-	MS^2	
$RFLa\text{-}PP\text{-}1^{47-73}$	YVEAADDDDVDAAERDNVIGTGAQPIW-OH	2905.31	-	MS^2	
RYamide (RYa)					
RYa^{7-38}	VYEEEPISGGRQLTGGVQAVAQPESFALGSRY-NH_2	3394.70	-	MS^2	
PP-1	NGNADDGTMNVAT-OH	1279.52	-	MS^2	
short Neuropeptide F (sNPF)					
sNPF-1	SNRSPSLRLRF-NH_2	1331.77	MS^2	MS^2	(Clynen et al. 2009, Dillen et al. 1097)
$sNPF\text{-}1^{4-11}$	SPSLRLRF-NH_2	974.59	MS^1	MS^2	(Clynen et al., 2009, Dillen et al. 2014)
sNPF-2	SDPLFGAPSAAAGSGQDSLAVAARSPSLRLRF-NH_2	3171.67	MS^1	MS^2	

PP_a-1	APSYPDYDNVRDLYELLLQREAEGARLAAAADDHQLV-OH	4158.05	MS^1	MS^2	
PP-2	SDPLLSNQLGAPESPVEN-OH	1866.91	-	MS^2	
SIFamide (SIFa)					
SIFa	AAATFRRPPFNGSIF-NH_2	1650.89	MS^1	MS^2	Gellerer et al. 2015
Sulfakinin (SK)					
SK-1 (pQ, SO_3)	pQLASDDY(SO_3)GHMRF-NH_2	1501.58	-	MS^2	Clynen & Schoofs 2009
SK-1 (pQ)	pQLASDDYGHMRF-NH_2	1421.62	MS^1	MS^2	
SK-1 (Q, SO_3)	QLASDDY(SO_3)GHMRF-NH_2	1518.61	-	MS^2	
SK-2 (pQ, SO_3)	pQPAAAPAAAAAPVPVAPRFDDY(SO_3)GHFRF-NH_2	2872.36	-	MS^2	
SK-2 (pQ)	pQPAAAPAAAAAPVPVAPRFDDYGHFRF-NH_2	2792.41	MS^1	-	
SK-2^{19-27} (SO_3)	FDDY(SO_3)GHFRF-NH_2	1282.49	-	MS^2	
Tachykinin-related peptide (TKRP)					
TKRP-1	APLLGFHGVR-NH_2	1065.63	MS^1	MS^2	
TKRP-2, 8	APSLGFHGVR-NH_2	1039.58	MS^1	MS^2	(Clynen & Schoofs 2009)
TKRP-3	APLRGFQGVR-NH_2	1099.65	MS^1	MS^2	
TKRP-4	ALKGFFGTR-NH_2	995.58	MS^1	MS^2	
TKRP-5	APQAGFYGVR-NH_2	1064.56	MS^1	MS^2	(Clynen & Schoofs 2009)
TKRP-6	GPSGFYGVR-NH_2	938.48	MS^1	-	Clynen & Schoofs 2009
TKRP-7	APLSGFYGVR-NH_2	1065.58	MS^1	-	(Clynen & Schoofs 2009)
TKRP-5+6+7	ALKGFFGTRGKKAPQAGFYGVRGKKGPSGFYGVR-NH_2	3588.98	-	MS^2	
TKRP-9	APVGFYGTR-NH_2	966.52	MS^1	MS^2	
ext. TKRP-9	GNTKKAPVGFYGTR-NH_2	1494.82	-	MS^2	Veelaert et al. 1999
PP-1	pQGPDAGEQRGPADAAAFLRMRAAGGDGDGKDALLE-OH	3466.64	MS^1	MS^2	
TKRP-1 + PP-2	APLLGFHGVRGKKDDLDELD-OH	2195.15	-	MS^2	
TKRP-2 + PP-3	APSLGFHGVRGKKDDADDDDGFD-OH	2434.09	-	MS^2	
PP-5	DDGASPPDLDSLLYYLNEASEATRQ-OH	2740.26	MS^1	-	
TKRP-8 + PP-5^{1-11}	APSLGFHGVRGKKDDGASPPDLDSLL-OH	2649.36	-	MS^2	
PP-5^{12-23}	YYLNEASEATRQ-OH	1444.67	-	MS^2	
PP-6	SWAPDGGAASSSDSIAPSLINSQ-OH	2218.03	-	MS^2	
Tryptopyrokinin 1 (tryptoPK1)					
tryptoPK1-1	EAAIHEDPGVWFGPRY-NH_2	1842.89	MS^2	MS^2	
tryptoPK1-2	AAKSPALWFGPRV-NH_2	1398.80	MS^2	-	
tryptoPK1-3	AQPPGLWFGPRV-NH_2	1323.73	MS^2	-	
tryptoPK1-4	SDAEVDDMLWFGPRP-NH_2	1733.80	MS^2	MS^2	
tryptoPK1-5	AAKHPGLWFGPRF-NH_2	1482.81	MS^1	-	

PP-3	SSDRQPKGEDALWTDDEREFKDGGSQRQD-OH	3352.51	MS^1	-	
PP-4	SVDVDKQEQTEDMYEDGVATRDQ-OH	2658.15	MS^1	-	
Tryptopyrokinin 2 (tryptoPK2)					
tryptoPK2-1	SLPEPGTWFGPRV-NH_2	1441.76	MS^2		
tryptoPK2-2	SNPDPGMWFGPRV-NH_2	1458.69	MS^1		
tryptoPK2-3, 6, 8, 9	SHPEPGMWFGPRV-NH_2	1495.73	MS^2		
tryptoPK2-4	SHPEPAMWFGPRI-NH_2	1523.77	MS^1		
tryptoPK2-5	SSPEPGMWFGPRV-NH_2	1445.70	MS^1		
tryptoPK2-7	SRPEPGTWFGPRV-NH_2	1484.78	MS^1		
tryptoPK2-10	SYPEPGMWFGPRV-NH_2	1521.73	MS^1		
PP-1	LDSIERSGGIRDSETRKRSENFSRS-OH	2882.45	MS^1		
PP-3	SVGFRVDHSYAKPGPHI-NH_2	1865.98	MS^1		

Table 2. Neuropeptides, additional precursor peptides (PPs), and peptides from neuropeptide-like precursors of *S. gregaria*, identified by mass spectrometry (direct tissue profiling with MALDI-TOF MS including MS^2, extract analysis with Q-Exactive Orbitrap MS). Distinct peptides from different transcripts are marked with subscripts (e.g., $NPF1_a$, $NPF1_b$). Numbering of peptides according to their position in the precursor. Neuropeptides originally described from *S. gregaria* with other names are marked with asterisks, references referring to gene/precursor description without confirmation of mature peptides are given in parentheses. Half of disulfide bridges underlined. APRP, Adipokinetic hormone-related peptide.

Supporting information S1: List of neuropeptide and neuropeptide-like precursors from *S. gregaria*. Blue, signal peptide (if incomplete or with low probability, the phrase "hypothetical signal peptide" is added to the header line); yellow, predicted sequence of bioactive neuropeptide; green, predicted C-terminal glycine amidation site of neuropeptide; red, cleavage sites of neuropeptide (confirmed cleavage sites in neuropeptide-like precursors); light grey, predicted C-bridge site; red letters, amino acids substitutions in alleles. If different alleles or transcripts are listed, confirmation of identical sequences by mass spectrometry is included only in the first precursor sequence.

MASS SPECTROMETRY MS^2
MASS SPECTROMETRY MS^1

> *Adipokinetic hormone 1*
MVQRCLVVALLVVVVAAALCSAQLNFTPNWGTGKRDAADFGDPYSFLYRLIQAEARKMSGCSN

> *Adipokinetic hormone 2*
MRQSCALTLMLVVAVCAALSAAQLNFSTGWGRRYADPNADPMAFLYKLIQIEARKLSGCSN

> *Adipokinetic hormone 3* (ortholog to *L. migratoria* AKH4)
MRGAGVLAVLLAALVAAGCCALCSAQLTFTPSWGKRAPPEAGAAYPSAEPFLYLYKLIQAEAQKMAGCSKFPN

> *Adipokinetic hormone/ corazonin-related peptide* (hypertrehalosemic hormone in Siegert, 1999)
MVARLFLALTVTAWCCYLVTSQVTFSRDWSPGKRSPEPTCAKHAASICQILVNELRQLAACEMKSLLRYHAEEVN
VPQEIYIDGNGGR

> *Agatoxin-like peptide* transcript a
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSDGNLERLLQGAQQKRSSFIYLFRRACIRRGGTCDHR
PKDCCYNSSCRCNLWGANCRCQRMGLFQKWGK

>*Agatoxin-like peptide* transcript b
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSDGNLERLLQGAQQKRACIRRGGTCDHRPKDCCYNSS
CRCNLWGANCRCQRMGLFQKWGK

> *Agatoxin-like peptide* transcript c
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSDGNLERLLQGAQQKEPESFSYGGEVSKREPARRACI
RRGGTCDHRPKDCCYNSSCRCNLWGANCRCQRMGLFQKWGK

> *Allatostatin A*
MGMTSRSSSSEAARLPLPALVLLLLCTTPATPQEVPGDAMTGGGPASAPVSTASEAAAASPPGSASTGAAPMDAE
SEYDLYKRLYDFGVGKRAYTYVSEYKRLPVYNFGLGKRATGAASLYSFGLGKRGPRTYSFGLGKRGDDEPNDYSE
QELFADVDGDSEDALPVAVEADERELPEAAEEEMPGVFTELMDKRGRLYSFGLGKRARPYSFGLGKRAGPAPSRL
YSFGLGKREGRMYSFGLGKRPLYGGDRRFSFGLGKRAPAEHRFSFGLGKRDARSADSQ

>*Allatostatin C* (allele 1)
MVSVAEVTLLFLAGSSGFAMPVDDVEEGYAGLRPCASSICLLLRPYIVHTDVPETEANRYIHKRQRKPRYYRCYF
NPISCF

> *Allatostatin C* (allele 2)
MVSVAEVTLLFLAGSSGFAMPVDDVEEGYAGLRPCASSICLLLRPYIVHTDVPETEVNRYIHKRQRKPRYYRCYF
NPISCF

> *Allatostatin CC*
MARHRSSAPTQSESAALLLLLALPALLCVAGAAPAPAPPAADAISVESRAIRPPIPQPPDYQDYQSAVRYDEYPV
VVPKRTALLLDRLMVDLKHLMDKDRGEAQNPIDSGSSIGRMALQRRGQKTGQYWRCYFNAVTCFRRK

> *Allatostatin CCC* (AstC in Veenstra 2014)
MAMSTAVKAVLLLVVALAATCWARAEPLGQQPSDKARLLNELDLVDDDGSIETALINYLFAKQVVNRLRAQMDVS
DLQRKRSYWKQCAFNAVSCFGK

> *Allatotropin*

MRCAAAALCLLVALAALCAAAAAAPAAHYGRGSRPRTIRGFKNVALSTARGFGKRDGNQLEAALADRDTTLPDSF
PVEWFAAEMQNNPELARMIVSKFVDANQDGELTAEELLRPTY

> *Bursicon alpha* (complemented with GHHP01009596.1)

...*AAPTPAPAV*VRPDASEASQAPERSPADECQLTPVIHVLQYPGCVPKPIPSFACTGRCSSYLQVSGSKIWQME
RSCMCCQESGEREASVSLFCPKAKPGERKFRKVSTKAPLECMCRPCTGVEESAVIPQEVAGYPDDGPLAAHFRKS
Q

> *Bursicon beta* (complemented with GHHP01011612.1)

MQAPSTWSSVVPLSLVILAAALAPVARPQPAGGGEEACETLPSEIHIIKEEFDELGRLQRTCTSEVGVNKCEGAC
NSQVQPSVTTPTGFLKECYCCRESFLRERTVTLSHCYDPDGARLTAEGTATMDIRLREPAECKCFKCGDFSR

> *Calcitonin A* (underlined sequence missing in Veenstra, 2021)

MTSTVGVLTAMVAAIALATGSWVPPDGDSARDVENSSILEPVFIRRLTIDGLRHRDYGTRGKRRSGVVCTDVAGE
PRRCFYEELVEMRRPEDVLNDLLSAKRDRQPLESHDVGKRRRTLECYIGGRMGGCDYQDIKQAQGEDQHLNSIDS
PGKRDLD

> *Calcitonin B* (allele 1)

MMKVVVLLCLFAVGLANPTEQLQVAAARGAAEVHRAKRCANLWDDSCLNDVSGASDDGDYFGSGNSPGKRWAPQQ
VRHLLEQMGRQAAAKRCANLWDDSCANGGFIGASEDGQYFGSGNSPGKRALSDQLRRLQLLRARAGSKRCANLWD
DSCANGGVGGASDDGHYFDSDQSPGK

> *Calcitonin B* (allele 2)

MMKVVVLLCLFAVGLANPTEQLQVAAARGAAEVHRAKRCANLWDDSCLNDVSGASDDGDYFGSGNSPGKRWAPQQ
VRHLLEQMGRQAAAKRCANLWDDSCANGGFIGASEDGQYFGSGNSPGKRALSEQLRRLQLLRARAGSKRCANLWD
DSCANGGVGGASDDGHYFDSDQSPGK

> *Calcitonin-like diuretic hormone (DH31)*

MHLASLVTSLLAVVLLLVTPARPAWANQLSNNYVSDYEMEQTAPLLSILELISKLRQTSSIAEDPAAKRGLDLGI
NRGFSGAQAAKHLMGLAAAQYAAGPGRRRRDTHQQPQQPAGHTSA

> *Capa* transcript a (allele 1)

MPARRYTAAALLAALALAAAAAAAAAEDGDKGISKLKKTSSLFPHPRIGRSEFINHEAGSGGQKRAAGLFQFPRV
GRAFIDAPLPFPLSVVGPLRLAPSDKDADAEAATVAEQPSPFEGHKRKGLVANARVGRRDGAETPGAAASLWFGP
RVGRAGLGQDETRAGTKRRGLLAFPRVGRGHAGSSSSSSSGDGDGARDSLWFGPRVGRRERRSLRLRLPAAAWLA
AGDVGNGKGDFTPRLGRESGEEEATVLLVGDGNTAEGFDAVADADIDEEER

> *Capa* transcript a (allele 2)

MPARRYTAAALLAALALAAAAAAAAAEDGDKGISKLKKTSSLFPHPRIGRSEFINHEAGSGGQKRAAGLFQFPRV
GRAFIDAPLPFPLSVVGPLRLAPSDKDADAEAATVAEQPSPFEGHKRKGLVANARVGRRDGADTPGAAASLWFGP
RVGRAGLGQDETRAGTKRRGLLAFPRVGRGHAGSSSSSSSGDGDGARDSLWFGPRVGRRERRSLRLRLPAAAWLA
AGDVGNGKGDFTPRLGRESGEEEATVLLVGDGNTAEGFDAVADADIDEEER

> *Capa* transcript a (allele 3)

MPARRYTAAALLAAFALAAAAAAAEDGDKGISKLKKTSSLFPHPRIGRSEFINHEAGSGGQKRAAGLFQFPRVGR
AFIDAPLPFPLSVVGPLRLAPSDKDADAEAATVAEQPSPFEGHKRKGLVANARVGRRDGAETPGAAASLWFGPRV
GRAGLGQDETRAGTKRRGLLAFPRVGRGHAGSSSSSSSGDGDGDGARDSLWFGPRVGRRERRSLRLRLPAAAWLA
AGDVGNGKGDFTPRLGRESGEEEATVLLVGDGNTAEGFDAVADADIDEEER

> *Capa* transcript b

MPARRYTAAALLAALALAAAAAAAAAEDGDKGISKLKKTSSLFPHPRIGRSEFINHEAVAEQPSPFEGHKRKGLV
ANARVGRRDGAETPGAAASLWFGPRVGRAGLGQDETRAGTKRRGLLAFPRVGRGHAGSSSSSSSGDGDGARDSLW
FGPRVGRRERRSLRLRLPAAAWLAAGDVGNGKGDFTPRLGRESGEEEATVLLVGDGNTAEGFDAVADADIDEEER

> *Carausius neuropeptide-like precursor 1* (allele 1)

*MKLCVAVLLVAAGVAATLHA*TGGEPVAGGVQYQGPARLKRDPLGFNARGFHDDIFGQEFGVFHTVKRSRSVESVE
RPLMVSVSESSGKNGRYKREVGALPSMPSVTKAMEYSKRRPEMGSSGFHGDTFSSGFGEFWTMKKKGHVKRDPEV
AEPIPYSDIWSDEEVEKKSGINEEFASDELPARFREEKLLFGQQLSDGARRQLLKNKGDWQRIISASFPVKPYLR
GNEDAVLDMLRSRMGSRKSERDPGFDDAAEQRSFFERSPQVGETVINRDALSNGLADLWKTSGNSQQENLQKRRP

EMGSSGFHGNMFSNGFGEFWPMKKKSNTLDMDGDVGVDKLLPMKRRPEMDSLGFHGDTFHNGFGDFWPMKRSTPE
ENFATGRSDPKTEQNYTHSKECCGDNEKN

> *Carausius neuropeptide-like precursor 1* (allele 2)
*MKLCVAVL*LVAAGVAATLHATGGEPVAGGVQYQGPARLKRDPLGFNARGFHDDIFGQEFGVFHTVKRSRSVESVE
RPLMVSVSESSGKNGRYKREVGALPSMPSVTKAMEYSKRRPEMGSSGFHGDTFSSGFGEFWTMKKKGHVKRDPEV
AEPIPYSDIWSDEEVEKKSGINEEFTSDELPARFREEKLLFGQQLSDGARRQLLKNKGDWQRIISASFPVKPYLR
GNEDAVLDMLRSRMGSRKSERDPGFDDAAEQRSFFERSPQVGETVINRDALPNGLADLWKTSGNSQQENLQKRRP
EMGSSGFHGNMFSNGFGEFWPMKKKSNTLDMDGDVGVDKLLPMKRRPEMDSLGFHGDTFHNGFGDFWPMKRSTPE
ENFATGRSDPKTEQNYTHSKECCGDNEKN

> *CCHamide 1*
MSRRSHSGPLAALAVAALAVAFLCAAAEGSCLSYGHSCWGAHGKRSGGGGASPRLGLVPDDARWTVSRLVA

> *CCHamide 2* transcript a
MSAKQIPAQPGCARLPTMALVLALAVTLALLQAADAKRGCMAFGHSCFGGHGKRADMEPAAEGVEGAEEAAAVAE
AEAAAAAALLDDAASPQRFRLSPFLRQWLQRAYQQQSADSQTVEVK

> *CCHamide 2* transcript b
MSAKQIPAQPGCARLPTMALVLALAVTLALLQAADAKRSYGVRRPGDIQVRRAGGCMAFGHSCFGGHGKRADMEP
AAEGVEGAEEAAAVAEAEAAAAAALLDDAASPQRFRLSPFLRQWLQRAYQQQSADSQTVEVK

> *CNMamide*
MRRRSALGSAALLLAAGLWLAGPGPSPVAARAMLLPLPRDAARFAAAAAAAEAEYDAVAGANWPDDEPPPQGLSQ
EEQTRLLRMLLLRMMQERDQHAQADHDAGPLLEPGDVADDPEGDGEHPLPYGGPHTPDDKRNLQGMMPTLCHFKI
CHMGRKRNQQKGARTAH

> *Corazonin*
MMRPWVSVVLLLVACWCLGALVHCQTFQYSHGWTNGRKRAGSSSAPGALLPPGRLPPPAAASDMDAQPCRVRCLR
LLLQGGAVPQLYVPPELWQQVDEEGDNMAARQRGGGARLRHALPPPGAAAAVDSDEDM

> *Corticotropin-releasing factor-like diuretic hormone*
MSPVRVLVAALLAVSCGGGCSAYYEAPPDGQRLLLQAAPAAAPAAASWPHQQRRQAIDEFAAAAAAAADAQYQDE
EEDGARRVKRMGMGPSLSIVNPMDVLRQRLLLEIARRRLRDAEEQIKANKDFLQQIGKRSPHAGGAANDAGADAP
PFGLRAAAERSASDISKDWASSDSRWNNQFTVRQS

> *Crustacean cardioactive peptide*
MSRALLMLCPALLVLLACLQRAAADDVIMEKRDMDSPFLNRLFEAKMKRPFCNAFTGCGKKRSDESVGTLLEMNS
EPAVADLSRQILSEAKLWEAIQEARAELLARRRQHEMQTNRLADFSRPLAVAQYRKKRAAAPPTPVQGMKPWRR

> *Ecdysis-triggering hormone*
MLLFKETFASLAVLVVLVAAAAAEPDEGANLFLKASRSVPHVGRRSDFFLKTAKSVPRIGRRSDLFLKSAKSVPR
IGRRTNLAAIEAQDGSEWLWPGGADAMPAPVRRQAYYVRKDGQPVMWSDVARDVEENPDLWPWNDFDAGNTREVD
NSR

> *Eclosion hormone 1*
MALCRPLLLTLALVALSVLVAAVYLTPGAAASAVGVCIRNCAQCKKMFGPYFEGQLCGDACLKFKGKMVPDCEDA
ASIAPFLSKLE

> *Eclosion hormone 2* (allele 1)
MAARRCCPLATLLAVLLVTSCLAPPADANAVSVCIHNCAQCKKMYGPYFEGQLCADACLKFNGKMMPDCEDAASI
APFLNKLE

> *Eclosion hormone 2* (allele 2)
MAARRCCPLATLLAVLLVTSCLAPPADANAVSVCIHNCAQCKKMFGPYFEGQLCADACLKFNGKMMPDCEDAASI
APFLNKLE

> *EFLamide*
…VLLYLFSSIYIYIYRYFNCYTFSVTEFLKKLSLWNRALPEMQQSTVYNNLVQFRVIRNLGSEFLGKRMRSLNNL
LGLK

> *Elevenin*
MTKTHQLSLLMITLCALCLAIGSAKGQKRVDCRMYPFAPICRGIMTKKRDVDSTGIASATIQKSADYPDGRQWSA
PEQAAAMFPWLLNNWPKRTAADYDY

> *Extended FMRFamide*
MTPPALLLLLLLMAAGGALQSRADADAEAAAVDDDGAGGEGDGDGELGLVQTTPRSNFLRLGRAGGAHSAFLRLG
RDRASSGFLRLGRGSERNFLRFGRSRQDGEQVAEDDELLSRAGRSGMADPLTRHDRNFIRFGRSGRSGTAQHQAG
AEPALWAALLDPRLAVLPAPAGADDADVDGAAGSGRGRVSTRSFLRLG

> *Glycoprotein hormone alpha 2*
MVPPSSRSALHFFALAVALCLSAVSAGMDGERDAWEKPGCHRVGHTRKISIPDCIEFPITTNACRGFCESWSVPS
ALNTLRVNPHQAITSIGQCCNIMETEDVEVRVMCLDGPRDLVFKSAKSCQCYHCKKD

> *Glycoprotein hormone beta 5*
MAHCRLMYVCSALLAAAAAAAAVADAAMDPASTLECHRRLYTYKVTKTDADGRACWDVINVMSCWGRCDSNEISD
WRFPYKRSFHPVCLHDARERRSVRLRNCEEGAAPGTERYEFLEAVSCRCAICRSSEASCEGLRYRGQRSGPFKAL
GRR

> *Gonadulin* (hypothetical signal peptide)
MFGSRCSTHYAVGDMQLHVAVLVCVGVLCATAAASTCDGESILKLMRDACVVRRRRDASPGERAAGEEQLVLPRA
GRLPHGLHDQFWESLLELADSEEKTVLSRALRKSSKFHQLISVCCRRTCTAKDFRVLCGPPRKT

> *HanSolin*
MLCWLLVLVAPAVLAAPYVRDESAVAWGRPPGAEAWQQRLPPGHKADDDDVHTAEKRAMLAMLSRWRPLTSAVYQ
LLRPRTPRPASQEALSAETRGSRRPLGQPLRWGR

> *IDL containing*
MVRACAQLALTAATLVAFCAALPQTVMAIDLSRLYGHLSAKRNGEACHPYEPFKCPGGEGKGICISIQYLCDGAP
DCPDGYDEDTRLCTAAKRPPVEETASFLQSLLASHGPNYLEKLFGNKARDALAPLGGVEKVAIALSESQTIEDFG
AALHLMRSDLEHLRSVFMAVENGDLGMLKSLGIKDSELGDVKFFLEKLVNTGFLD

> *Inotocin/ (Vasopressin-like)* (hypothetical signal peptide)
MRFQQAVLLTVAVSIAAACLITNCPRGGKRAGLQHRCAPCGPGGQGVCLGPHICCGPRMGCRLASPADTAVCRST
APCPLDSPEMRCAGGVGRCSAPGVCCFKDSCHIDTSCVADTTESDDRYPVDTVVVSDDSCHIDTSCVADTTESAD
RYPVDTVAVSDGDREEMKI

> *Insect parathyroid hormone*
MSRIVFSLLLAAVLAAVAVSDGSAMPADNHRVKRMSDHRVAELQTLLSMGKMGGKVVTHKVGYGQVDPMKVGRRR
RSDPRMLSRLQQLLLSAAAGDSHEQLPQQEEEQQHPQLQLQLDPTEEDERQVLLQLLEGQRQAQSPQEQQQARLA
WWPRTLTPQL

> *Insulin-like growth factor*
MRPLLAVLCVACAGAFLCPRAAAEQRLRVCGRELAETLSLLCRDRGGFNDPPPPHQRVARRGGVADDCCRLGCSL
STLLRYCKFDDVVQPPRAHDDGDDACVVDDDDDDELFSLQLLHGPEAWCGTWQPAKPAPPSGEVECTCSRGGGGE
QLENVVRKVPRRPQRQRPRQLPRQPARQPPRQPPTTPATRRPVVRGTATPYFAGRPVAVVLTPPHTQRQPATL

> *Insulin-related peptide*
MMWKLCLRLLAVLAVCLCTATQAQSDLFLLSPKRSGAPQPVARYCGEKLSNALKIVCRGNYNTMFKKASQDVSDA
ESEDNYWSQSADEEVEAPALPPYPVLARPSAGGLLTAAVFRRRTRGVFDECCRKSCSISELQTYCGRR

> *Ion transport peptide* transcript a (hypothetical signal peptide)
MHHQKQQQQQKQQGEAPCRHLQWRLSGVVLCVLVVASLVSTAASSPLDPHHLAKRSFFDIQCKGVYDKSIFARLD
RICEDCYNLFREPQLHSLCRKDCFTSDYFKGCIDVLLLQDDMDKIQSWIKQIHGAEPGV

> *Ion transport peptide* transcript b (hypothetical signal peptide)
MHHQKQQQQQKQQGEAPCRHLQWRLSGVVLCVLVVASLVSTAASSPLDPHHLAKRSFFDIQCKGVYDKSIFARLD
RICEDCYNLFREPQLHSLCRSDCFKSPYFKGCLQALLLIDEEEKFNQMVEILGKK

> *ITG-like* (allele 1)

MWLIGRLVQVVVVLALLNGGALGWGGLFNRFSPEMLSNLGYGGHGYGAYRSSQPLLQRFHNPVEVFQELQEDEEP
CYGKKCTSNEHCCPGTVCVDVDGIVGSCLFAYGLKQGELCRRDSDCETGLLCADSADGRTCQPPLTNRKQYSEDC
TMSSECDISKGLCCQLRRRHRQAPRKVCSYFKDPLICIGPVAADQVKEDIEHTAGEKRITGKVASFNHIRRK

> *ITG-like* (allele 2)
MWLIGRLVQVVVVLALLNGGALGWGGLFNRFSPEMLSNLGYGGHGYGAYRSSQPLLQRFHNPVEVFQELQEDEEP
CYGKKCTSNEHCCPGTVCVDVDGIVGSCLFAYGLKQGELCRRDSDCETGLLCADSADGRTCQPPLTNRKQYSEDC
TMSSECDISKGLCCQLQRRHRQAPRKVCSYFKDPLICIGPVAADQVKEDIEHTAGEKRITGKVASFNHIRRK

> *Kinin*
MRALTVLAVALAAATTGLAWGAERCCGDAEPWQGQARSKRSFSAWGGKRTGPDGVAGADVDDGYGDGDDQKRAFS
SWGGKRAAAGWQAEDKRSKQSFYSWGGKRGVGGDAEADGQAPEDAVEKKRGGVRFSSWGGKRDPFQDDDGGVEDL
ADSKRAFSSWGGKRFGDGASAAEKRAFSNWGESLANRRLDLYNYRPVQLRNGNGPPFFPWGG

> *Myoinhibitory peptide* transcript a (= *Allatostatin B)*
MRHGLRPRAATASLLLLALLAALASCSLAAPQAPAADTVSAAAPAAAAASPAASDEDKSDEEKRAWQDLNAGWGK
RAWQDLNAGWGKRGWRDLQSAWGKRAWSNLHGAWGKRGGDDAATWPELPDQTVAEQEGDQESAQLVPLPLQVPLP
LQLQLQLQGDEGEPGDEEKRAWSSLHNTWGKRAADWRSFHGSWGKREPGWTNLKGLWGKRAGPSNWNRLPAMWGK
RSEEE

> *Myoinhibitory peptide* transcript b *(= Allatostatin B)*
MRHGLRPRAATASLLLLALLAALASCSLAAPQAPAADTVSAAAPAAAAASPAASDEDKSDEEKRAWQDLNAGWGG
ARGRGKTLGSSWGKRAWQDLNAGWGKRGWRDLQSAWGKRAWSNLHGAWGKRGGDDAATWPELPDQTVAEQEGDQE
SAQLVPLPLQVPLPLQLQLQLQGDEGEPGDEEKRAWSSLHNTWGKRAADWRSFHGSWGKREPGWTNLKGLWGKRA
GPSNWNRLPAMWGKRSEEE

> *Myosuppressin* transcript a
MRSCMVIVVALAVSAQLAWWGAQANAVPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAVRDPE
LFVEDKRPDVDHVFLRFGRRRR

> *Myosuppressin* transcript b
MRSCMVIVVALAVSAQLAWWGAQANAVPVTRPIVCATDDVSPQIRKVCQAYEAFSELATSAKDYLDHFAAEAHRQ
SVRPLPAGIKREDVGHVFLRFGKRR

> *Natalisin*
MPHAVAWAWLAAAATLALARQGPDSENATLAEDSARGRVKRAGVPEAGLADDGRSRRQEPLYVEEPAWVLLDRRE
TLGPARQAVAAAVAAPAGVGDAFWVARGRRREMSASSASAEGPFWAARGRREMEFFWPARGKRDQEEVMVGGSGE
EEDAGLAGEWEANRNRRAGLRSYLASMSGGQSFVPARGKRGSPHAGVSKRETGEVELASRRASLGSESVEGPFWA
PRGRRGQAARLLRLRLASAPAGSAFWPARGRRAGPSSASSSAEGPFWATRGRSLAAQHSFWPARGRRGDPTNDRG
DAAETEGEDSDNAEDRSFWEKLHEKDQARTHQ

> *Neuroparsin* transcript a
MKPAAALAAATLLIAVILFHRAEANPISRSCEGANCVVDLTRCEYGEVTDFFGRKVCAKGPGDKCGGPYELHGKC
GDGMDCRCGVCSGCSMQSLECFFFEGAAPNSC

> *Neuroparsin* transcript b
MKPAAALAAATLLIAVILFHRAEANPISRSCEGANCVVDLTRCEYGEVTDFFGRKVCAKGPSEHCNDFIKCGDGL
RCNCGRCTGCSMHTLQCYSDFSTPTTCP

> *Neuroparsin* transcript c
MKPAAALAAATLLIAVILFHRAEANPISRSCEGANCVVDLTRCEYGEVTDFFGRKVCAKGPGDDCNVYASCGIGM
DCRNKKCSGCSVKTLECYFESFPMSEEE

> *Neuroparsin* transcript d
MKPAAALVAATLLIAVILFHRAEANPISRSCEGANCVVDLTRCEYGEVTDFFGRKVCAKGPGERCSDYESCGDGL
RCAQKLCVGCSLATLQCFSIVSLPLSEEE

> *Neuropeptide F1* transcript a
MSQSRPLALLVLSAAVVALLLVVAAPAPAEAQQAAADGNKLEGLADALKYLQELDRYYSQVARPRFGKRSELRPD
VVDDVIPEETSADKFWRRFARRR

> *Neuropeptide F1* transcript b

MSQSRPLALLVLSAAVVALLLVVAAPAPAEAQQAAADGNKLEGLADALKYLQELDRYYSQVARPSPRSGGGASLS
VRSPLDTLNIAEHLRGVEKMVRMLQLQEYDRMYTPRNRPRFGKRSELRPDVVDDVIPEETSADKFWRRFARRR

> *Neuropeptide F2 (NPY)*

MSLVRVVVIALVVLAATAPLSEPLPAGADAGQQRPERPPMFTSPEELRNYLTQLSDFYASLGRPRFGKRGSAAAA
AFRAASRLPVPPPDAYEQLFQYDE

> *Neuropeptide-like precursor 1*

MRWCSLAALCAALAVAALRPPQAVAEPGDGLMVPGKRYVAALARNGELPLYGGWRKKQQRGDKRYTNRYGKRGAA
ADFETDDDEPESGVDALLEEIAATEQLRHLQLDALRRELDAQEAREEQAVLEAAEAEAEAEAEAAEAEGGGEEGA
DKRSVASLARAGALLPGTGKRNIAAMAKNGLLSPSGPVLLDGDGSGNGGEKRSVGALARGGLLPTPGRRAAAGDD
ADDDDLSLDSLMQQLYSEEEKRHIGSLARDYSLPSFGKRNLGSLARSGGLSNVRYVTTKKDDAERVGGHSEADKR
NIASFMRSRGSSLVEKRYLASLVRSHGLPYPLTKKEDDGPGEIKRNVGALARNWMLPSGKRASDDDQEVDKRYLA
SVLRQGRSDGFRQNSDGAQQADHEEEKRHLGSLAKSGMAIHKKTSRSAGSDGQAFLQQQQQQEQGGAHAQDAAGS
KRSKRHAYLLPPAPPQSLAPAPGEFPMPVLQNNDDALDYGDLLDLMSDVLGAPEKRFLGVPPAAADYGKRHIGAL
ARLGWLPSFRAASARSGRSAGSRSGKRATRSHSADGPWPAELQQA

> *NVP-like* transcript a

MNRGRRTRLGLPTVALLALAVVAGAVGAAALPTSLLEDASAATAHQQTDKGGEAPADAGSASTAAPGAAPETAGS
GGRASVSGDGKTVAQYEKGYLYGAGKAAPDKHVENALLKSELYGDPSAVNQYRYYGGANERRPDGAEGAFAPPSK
RSSSFRPMVPHALELSGVGPRLKRDLAVDPEDVLALLQLWQAERHGAPRVPSKWSRYGSVEGEEYPQAVGNENEE
MEEDDSNNGEWLEGPVYSSALGPHYAVDRRALYIPDYSYQVLNPFTAYQVQPDKRDSRWNGYSKDKRFMVSRKRD
VTQPARGDIHYLAQLLGPSHRDQQIPLFHRVAV

> *NVP-like* transcript b

MNRGRRTRLGLPTVALLALAVVAGAVGAAALPTSLLEDASAATAHQQTDKGAVNQYRYYGGANERRPDGAEGAFA
PPSKRSSSFRPMVPHALELSGVGPRLKRDLAVDPEDVLALLQLWQAERHGAPRVPSKWSRYGSVEGEEYPQAVGN
ENEEMEEDDSNNGEWLEGPVYSSALGPHYAVDRRALYIPDYSYQVLNPFTAYQVQPDKRDSRWNGYSKDKRFMVS
RKRDVTQPARGDIHYLAQLLGPSHRDQQIPLFHRVAV

> *Orcokinin* transcript a

MQTATLLVAALAAAAVALAGAVPAPQMVSSGFQQYRDEPNDVEEGLVRHLDNIGGGHLLRNLDGLGGGHLLRQTK
SGLDSLSGATFGEQKRLDSLSGITFGSQKRNFDEIDRSGFNSFIKKNFDEIDRSGFDRFVKKNFDEIDRSGFSGF
VKRNAPMLARHYDQGDH

> *Orcokinin* transcript b

MQTATLLVAALAAAAVALAGAVPAPQMVSSGFQQYRDEPNDVEEGLVRHLDNIGADYSVNELDTLKAIFQHSKKE
STSHHNSDSSGLDHKGSENLVRNIDSIGGGNLVRNIDPIGGGNLPRNIDSIGGGHLVRNIDSIGGGNLVRNIDSI
GGGNLVRNIDSLGGGNLVRNIDSIGGGNLVRNIDSIGGGNLVRTIDRIGGGHLIRNIDSIGGGNLPRNIDSIGGG
NLVRNIDSIGGGNLVRNIDSIGGGNLVRAIDRIGGGHLLRNIDSIGGGNLVRNIDSIGGGNLVRNIDRLGGGHLI
RNIDPIGGGNLPRNIDSIGGGNLVRNIDSIGGGNLVKRSE

> *Periplaneta neuropeptide-like precursor*

MQLRWILLLLVLPAAALAQARGDSLQAALDAVTRRQRDLTLPARPSGYYALSQYRHQAAPATDDDIAFLDTGRDF
TGYGQPENIGAGYQKTISSGAAPYPAVLSDVPVHPPPNQLENMLLDYMKDAYVKDGDEEEGNYYQYGMDTDAKRS
LFRERENDDSGQDVEYHQVLPTYYDSENKQENVLLPSLFRERFLTRPEIDQVVELENMRRYAAKAIAKQLETDEE
EKLENKRNNDEEEEYLSLLRNLWEKYKEAKPQLVDFDDLTQNDIQEILSSLRNDGSSLHKRQYGYGSGFDIFNNA
GLMGQWGTGANNFAKRNKQRVEGPGQQGANFLYSLKFVAPEVNREAVETLKDNEGIELPDERDEDVLRLASGFAR
NNPEELMQIYGRPATMDEIYSPNQETYQTLSLETPDIGSRATSTKHFSSLARDVNEYQQLSPPNYMVIPERTSNK
RFIYEAKRKRYPVTKRSSNFYASPPMLHHKSFNSEGIKDTNKKKSSPITGVTDPKVAQELNQIFSSSVTHDDSPK
ASEKKVDSGTNSKLTTSKDPKDPNHETTTTHSVTTVSVQENTQNNDTETGTKSAHNRTSGPREEGNHKSHAHGTS
LHMEKETAEPITMSKIQTPLEIKKKSIDWSDYFGIDRRRKKTAPETVETKDGDGNMIDDNWLLNQYYRTLAMVSN
PLKKRMVSHSHANSPNNNDNKRDTGKQQAYNTDIFSRSSQREATKKSLNTKEDTSIDDMDTKLRNMEDLIVNEAV
KYTGSHEGTQDPKEIQEMKDKIMSRLAAAYSLEKMRLALAEFKSSLQAQMMSKYNPANLKSSLSNDNTQEESKMK
RVAVKKEKAEDDKHNDDGDKKKKRNGIKNDAEEETSGEFLDGPVDVEPMSEGYMGRSSDDYETGCPILDQILQKC
RSIGYASEDHNQVFLSLCSLHQICNICGPEVGAPSSSACDLMFITEADSICEEEIKCQRMSHRILTLLHREGHVT
DGNQVIEAITPTEGHVINETEQGNEDFHHAAPEFHMPVYSERTTITIQPRRTDDDRVPMPVSSANWQTTSFTHFI
PLSGLDFLTVADQLIIQQTVELQDILATVESENRYLVKVPHGEILYYAAEASETLQRMFCGSSRGFYMKLFDQTR
QQAILLKRRLACANCCCGCYLQEMKLFTPGGVLLGSVEQEWTLINPSFVVRDSDGNIEYVIEGPQHFPCTNYDGT
TFKVLSRDGLTQLSTIVHEWNNLTGDFNIIVTFPTDNLHPKMKAVFLGAGFLLEYMFFETSKRTNRCQKWCR

> *Pigment dispersing factor*
MTIMAVSGKLLTAFVLSIYVLGLALTIHATQYEEEKYQENEVRYGRELATLLAQLAHRNEPAICAHKRNSEIINS
LLGLPKLLNDAGRK

> *Proctolin*
MFSQRKALVLVLVVMLVMSVAQARYLPTRRSSASLEDRLDRLRDLINDLVESERPSARMAPPRRLDWILAGEPEL
DYAQ

> *Prothoracicotropic hormone*
MAIFWTFLVLVSCAMASATSLEDRLRPLWTEAEAEAAAAALAGAAGAAGPVGACSDPEDCAFHRSGRGRGGHPRV
LEVAMPCHCRQEHSFQKLGVGHYPQLLRQVGCRRDAACCLEPYRCRQLNHSVLVLVDSHIVPDELDEYSAEKLPK
ELRDSHWNFKAVSVGVGCQCVMDVLAS

> *Pyrokinin*
...RAACRFAGGGGSGSGWVGRREGDFTPRLGRESAEQGAAPQWQSAEEQVLSGPFVPRLGRGAAPAAQFSPRLG
RRDPPADGLVWLPLVPRLGRRRPLPAPAAPFVPRLGRDSSEDWAQPFVPRLGRRLQQYGMPFSPRLGRDAPDQLQ
ADE

> *PK-like*
MAGAAAPAAAALAAVAVLAAASAPAATTHGTTDTGRARLAWDWLQASPRMRRDALGDMMLAHVIEHPWIVMPLA
AVGSAPVGAAPPPASRRQSMPTFTPRLGRDSAGDDLAEEEAEGDGDADGGHGQGQLQLAPPFWPRPGRHAPPLPL
TPRLGRAAHQATPAASGEPTSA

> *RFLamide*
MLLPLLLATLAAGPVLSEAAPAPAPQAAARSPTEITADPEQLELLSAYLTRLLVDGWPEGEPPLLYVEAADDDDV
DAAERDNVIGTGAQPIWKRSRYYRRYPWKRQNGRYQAADDRYLCQPTREDVFQLLVALHEARRGDSSRTVSFCNR
RRPASAIFTNIRFLGR

> *RYamide*
MTSTSNELSEMVARVLLLAFMLFAVAQAISLYDNVYEEEPISGGRQLTGGVQAVAQPESFALGSRYGKRNGNADD
GTMNVATRDADNSDSEDRQSVANYVVFLDNCLCSCKKPSGGLN

> *Salivary gland salivation stimulating peptide*
MAPKTVFALLVVVCLAELTWASPAPFHSRRSTNATTATATNATANSSSNDAASDDDSSSSNIFSEIKEHVDNVFG
NIFGKRSVDSSNSSDVSSSSNTTTTSNDTSSNNVFDEIKDNVEGFFGHLFGKRAVKFPTPDMFTKLYEEWANQRA
NRSVTVRATNSSFTMGDLFTEVVKGWLNGDNSKRSVTVRETDNSFTMGDLFVELFKEWLKGNLDKRSVTVRETDN
SFTMGDLFAELSKEWLKGNLDKRSVTVRETDLFTELVKEWLKGNLDKRSVTVRDTDNSFTMGDLFTELVKGWLNG
ENNKRSVTVRDTDNSFTMGDLFVELFKEWLKGNLDKRSVTV

> *short neuropeptide F* transcript a
MASTSAVCKVALVLLLVAALAASAPSYPDYDNVRDLYELLLQREAEGARLAAAADDHQLVRKSNRSPSLRLRFGR
RSDPLFGAPSAAAGSGQDSLAVAARSPSLRLRFGRRSDPLLSNQLGAPESPVEN

> *short neuropeptide F* transcript b
MASTSAVCKVALVLLLVAALAASAPSYPDYDSVADVRDLYELLLQREAEGARLAAAADDHQLVRKSNRSPSLRLR
FGRRSDPLFGAPSAAAGSGQDSLAVAARSPSLRLRFGRRSDPLLSNQLGAPESPVEN

> *SIFamide*
MQSAVYSRLLVAVVVALLVFAGTASAAAATFRRPPFNGSIFGKRNSVEYTGSSTAVSAVCEIAAEACAAWLNNEK

> *SMYamide*
MKRCSFSCMLVWAIVLQICLTEGIGFKKLPFNGAMYGKRTSSVDYDSSNRAISSLCETASELCSSWYSQPDSN

> *Sulfakinin*
MSGRAVLLMVAVAVGAWLSGAAPSASGAAAPGGARRLEELLLSAPYDDPALVDDLLEAASKRQLASDDYGHMRFG
KRQPAAAPAAAAAPVPVAPRFDDYGHFRFGKRRSQ

> *Tachykinin-related peptide*
MCRVGALLLMAALVACEGAAGQGPDAGEQRGPADAAAFLRMRAAGGDGDGKDALLEKRAPLLGFHGVRGKKDDLD
ELDKRAPSLGFHGVRGKKDDADDDDGFDKRAPLRGFQGVRGKKDEADAEDAELGDGADYLQLADLPYREDYDADA
DADADELQLDDPWLRDEKRALKGFFGTRGKKAPQAGFYGVRGKKGPSGFYGVRGKKAPLSGFYGVRGKKAPSLGF
HGVRGKKDDGASPPDLDSLLYYLNEASEATRQKRGNTKKAPVGFYGTRGKKSWAPDGGAASSSDSIAPSLINSQ

> *Trissin*

MTRYSIVTLLLAGATLVALCAATCDFCGRECSKACGTNYFRTCCLNYLRKRSGPPGLHLELMLLPPEAGAARASA
AADRLRGERERAAAAAAGRRAPGADGDADADAVQYIYSV

> *Tryptopyrokinin 1*

MPRGAQLFLLLALVTAARVLDTKAESALTENDSQSSSQDSRVASHDRNEETAEDRPDGEAPNASFRRGREAAGFD
SRIDGRNGIGDESRTDEGPWLAMLEGGYIPAGSAREVVQREAAIHEDPGVWFGPRYGRRSSCGDEAPLKWLPQVE
KRAAKSPALWFGPRVGRSSDRQPKGEDALWTDDEREFKDGGSQRQDRRAQPPGLWFGPRVGRRSDAEVDDMLWFG
PRPGRSVDVDKQEQTEDMYEDGVATRDQRAAKHPGLWFGPRFGR

> *Tryptopyrokinin 2*

MESSTDALAVALIILAALLLCVETSEVLPDSKEHNDGVSDSKIQVRSENFSRSKRNNSEVHVDKWFQPNIVDDED
YWDTLFGDGVEEHVPKSDASVANIIQTSETPHLSLMQPTSRRSVGFRVDHSYAKPGPHIGRSLPEPGTWFGPRVG
RSNPDPGMWFGPRVGRSHPEPGMWFGPRVGRSHPEPAMWFGPRIGRSSPEPGMWFGPRVGRSHPEPGMWFGPRVG
RSRPEPGTWFGPRVGRSHPEPGMWFGPRVGRSHPEPGMWFGPRVGRSYPEPGMWFGPRVGRSHSESELRM

Supporting information S2: Updated list of neuropeptide and neuropeptide-like precursors of *Locusta migratoria*, complementary to the data from Veenstra, 2014. Blue, signal peptide.; yellow, predicted sequence of bioactive neuropeptide; green, predicted C-terminal glycine amidation site of neuropeptide; red, cleavage sites of neuropeptide (confirmed cleavage sites in neuropeptide-like precursors); light grey, predicted C-bridge site.

> *Agatoxin-like peptide* transcript a (accession number *pending*)
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSEGNLERLLQGAQQKRSSFIYLFRRACIRRGGTCDHR
PKDCCYNSSCRCNLWGANCRCQRMGLFQKWGK

> *Agatoxin-like peptide* transcript b (accession number *pending*)
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSEGNLERLLQGAQQKRACIRRGGTCDHRPKDCCYNSS
CRCNLWGANCRCQRMGLFQKWGK

> *Agatoxin-like peptide* transcript c (accession number *pending*)
MRTSLALMLALAAILTTHLAAAGPYLDDPVPDDGVEDYSEGNLERLLQGAQQNEPESFNFAGEVSKREPARRSSF
IYLFRRACIRRGGTCDHRPKDCCYNSSCRCNLWGANCRCQRMGLFQKWGK

> *Allatostatin C* (accession number GBDZ01064753.1)
...TGYAGLRPCASSICLLLRPYIVHADDPEPDSNRFIYKRKRQLRYYRCYFNPISCF...

> *Calcitonin A* (Hou et al., 2015)
MTSMAGVLAAMVVVVALATGSWISSEGDAATDVENSSILEPVLIRRLTVDGARLRDRDARGKRPGGVVCTDVAGE
PRRCFYEEFVEVRRAGDVVSDLVPAKAERQPVASHDVGKRRRTLECYIGGRMGGCDYQDLKQAQGEDQHLNSIDS
PGKRDFD

> *Carausius neuropeptide-like precursor 1* (accession number *pending*)
MLLSVAVLLAAACAAANHHAAGGVAPADQYPGPARLKRDPLGFNARGFHDDIFGQDFGIFHTVKRSRSVENVERP
LLVGVADSSDKNGRYKREVVALPAMPSTTKAMEFSKRRPEMGSSGFHGDTFSSGFGEFWTMKKKAHVKRDPEVPE
PVQYADMWSDEEAVKKSLPTEGFASEELPSRFREEKVLFEPQVSEGGPTRKLLKNKGEWHRIIPATFRIKPYVRG
NDDTVIEMLRSRMQTRGDSSLRNAAEQGNSLYEANPEIGETVVQRTELPDGLADLWKTNGNSQEENLQKRRPEMG
SSGFHGNMFSNGFGEFWPMKKRSNTLAMYGDADIGKLPPMKRRPEMDSLGFHGDTFHNGFGDFWPMKRFASAENF
ATVRDEPKMAQNYTHFKECCEDDEKTDLKISVR

> *EFLamide* (Veenstra and Šimo, 2020)
...FLKTLSLLNERLSEMQQSTTYNDLVQFRVVRNLGSEFLGKRMENLGSEFLGKRMQNLKNILGVK

> *Gonadulin* (Veenstra et al., 2021)
MRARLTGIRLAVGNMQLYVAVLVFVGALCASAAANTCDGERILELMKNACVVRRKREASPAARQAVHDQISQLRA
QRLPQDLDDQFWDGLQELAGSEQKTVLARALRKSSKFHQLITACCRRACTAKDFRLLCGSPRKP

> *IDL-containing* (accession number *pending*)
MVRACAQLALTAATLAAFCAALPQTVMAIDLSRLYGHLSAKRNGEACHPYEPFKCPGGEGKGICISIQYLCDGAP
DCPDGYDEDTRLCTAAKRPPVEETASFLQSLLASHGPNYLEKLFGNKARDALAPLGGVEKVAIALSESQTIEDFG
AALHLMRSDLEHLRSVFMAVENGDLGMLKSLGIKDSELGDVKFFLEKLVNTGFLD

> *Insulin-like growth factor* (Veenstra et al., 2021)
MRVLLAVTVFCACAVWLPAAADEERRYCGKYLAEKLWELCRHRGGFNEPPQTRPRSASQGGVARDCCRKGCSRQT
LLSYCKSDNVLKPQPPPPRPTEACSDDDEDGDELFSLPPLHREDWCGTWQPAAEPPPGGEVECTCSRGGGQQPDS
AVRKSPQPPTRQPRRQPPRQPPTTPASTPAPATRRPVVRGTVTPYFQGRPVVLSPPHAQRLPTSA

> *ITG-like* (accession number *pending*)
MWLVARLVQVGVALSLLSGGALGWGGLFNRFSPEMLSNLGYGGHGYGAYRSSQPLLQRFHNPVEVFQELQEDEEP
CYGKKCTSNEHCCPGTVCVDVDGIVGSCLFAYGLKQGELCRRDSDCETGLLCADSADGRSCQPPLTNRKQYSEDC
TMSSECDISKGLCCQLQRRHRQAPRKVCSYFKDPLICIGPVAADQVKDEIEHTAGEKRITGKVASFNHIRRK

> *NVP-like* transcript a (partial sequence under baratin in Veenstra, 2014) complemented with (accession number *pending*)

MNRIAVMGEELGGGLCAAALVALCLAARVAALPTSLLEDASAVANNPHVDTGATEGVLASTAPPGAAGAAAAGGG
ARGSGAGAVDGKTVAQYEKGYLYGAGKAAPDQHVENALLKSELYGDPSAVNQYRYYGGSNERRPDGAEGAFAPPS
KRSSSFRPMVPHALELSGVGPRLKRDLGVDPEDVLALLQLWQAERHAANRAPSKWSRYGNIEGEEYPQAVGNENE
EMEEDDSNNGEWLEGPVYSSALGPHYAVDRRALYIPEYPYQVVGPYAGYQVQPDKRDSRWNGFSKEKRFMVSRKR
DMSQPARGDIHYLAQLLGPSHRDPQNPLYHRVAA

> *NVP-like* transcript b (accession number *pending*)

MNRIAVMGEELGGGLCAAALVALCLAARVAALPTSLLEDASAVANNPHVDTGAVNQYRYYGGSNERRPDGAEGAF
APPSKRSSSFRPMVPHALELSGVGPRLKRDLGVDPEDVLALLQLWQAERHAANRAPSKWSRYGNIEGEEYPQAVG
NENEEMEEDDSNNGEWLEGPVYSSALGPHYAVDRRALYIPEYPYQVVGPYAGYQVQPDKRDSRWNGFSKEKRFMV
SRKRDMSQPARGDIHYLAQLLGPSHRDPQNPLYHRVAA

> *PK-like* (Redeker et al., 2017)

MLPLPALWASVLAAAGSFPAGAASASAWRRQSVPTFTPRLGRDSAGDELAEEDAVDDGGDGLPQPQLAPPFWPRP
GRRAPPTRGAAPPTS

> *Periplaneta neuropeptide-like precursor* (partial sequence in Zeng et al., 2021) complemented with (accession number *pending*)

MQLRWIVLLLVLPAAALAQANGDSLQTALDAVTRRQRDLTLPLPARPSGYYALSQYRHQATPPDDDDDDIAFLDT
GRDFTGYGQPENIGAGYQKTISSGAAPYPAVLSEVPVHPPPNQLENVLLDYVKDAYVKDDDDDEEGNYYQYAMDT
DAKRSLFRERENDDNGQDVEYRQLLPGYYDNENKQENILLPSLFRERFHTQPDIDQVVEMDKMRRYATNAVSKQL
EADEQEKAESKRNNDEEEEYLALLRNLWEKYKEAKPQLLDFDDLTQNDIQEILSSLRNDRSNIRKRQYGYGSGFD
IFNNAGLMSQWGTGSSNFAKRNKQRVEGAGQQGANFLYSLKFVAPEVNREAVETLTDNEGIELPDERDEDILHLA
SGFARNNPEQFMQIYGRPASVDEIYSPNQETYQTLSLETPDIGSRATSTKHLSPLPRDASEYQQLPPANYMVIPE
RSSNKRFIYEAKRKRYPVTKRSSNFYASPPMLHHKSFNSEGIKDVNKKKSTPITGVTDPKVAQELNQIFSSSVTH
EDSSKDNEKKLDSGNNAKSTTSKETKDSVHETTTSHSVTTASVQENTLNNDTETKANFAHNRSGSGEEGSHKSSV
HGSSLHMEKETAEPITMSKIQTPLEIKKKSIDWSDYFGIDRRRKKTVPENVETKDAEGNMIDDNWLLNQYYRTLA
MVSNPLKKRVASHSHSYSPDNNDNKRDNGKPQAHNPDIFSRSAQRETVKKNMNIKEDSSIDEMDTKLRNMEDLIV
NEAVKYTGSHEGTQDPKEIQDMKDKIMSRLAAAYSLEKMRLALAEFKSSLQAQMMSKYNPANLKSTSSDSNAQEE
SKMKRVAVKKEKAEDDKHDDTDKKKKRNGNKNDVEEETSGEFLDSPVDVEPMSEGYMGRSVDDYETGCPILDQIL
QRCRSIGYASEDRNQAFLSLCSLHQICNICGPEVGAPTSSACDLMFITEADSVCEEEIKCQRMSHRILTLLHRGH
PLGISGISTDHCSNDSCLAQYFLTSPLPNNQPR

References

Hou, L., Jiang, F., Yang, P., Wang, X., Kang, L., 2015. Molecular characterization and expression profiles of neuropeptide precursors in the migratory locust. Insect Biochemistry and Molecular Biology 63, 63–71.

Redeker, J., Bläser, M., Neupert, S., Predel, R., 2017. Identification and distribution of products from novel tryptopyrokinin genes in the locust, *Locusta migratoria*. Biochemical and Biophysical Research Communications 486, 70–75.

Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Frontiers in Physiology 5, 1–22.

Veenstra, J.A., 2016. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. General and comparative endocrinology 230-231, 153-157.

Veenstra, J.A., Leyria, J., Orchard, I., Lange, A.B., 2021. Identification of Gonadulin and Insulin-Like Growth Factor From Migratory Locusts and Their Importance in Reproduction in *Locusta migratoria*. Frontiers in Endocrinology.

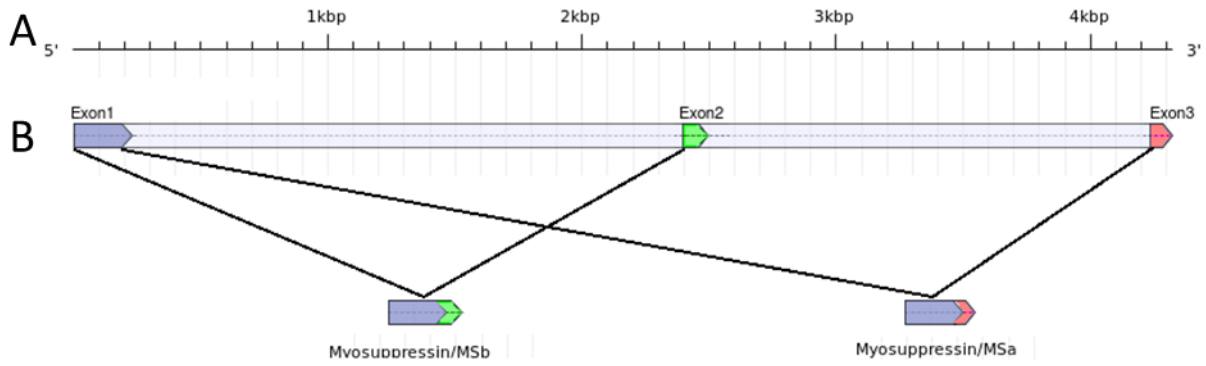
Veenstra, J.A., 2021. Identification of cells expressing Calcitonins A and B, PDF and ACP in Locusta migratoria using cross-reacting antisera and in situ hybridization. bioRxiv https://doi.org/10.1101/2021.07.28.454216

Veenstra, J.A., Šimo, L., 2020. The TRH-ortholog EFLamide in the migratory locust. Insect Biochemistry and Molecular Biology 116, 103281.

Zeng, H., Qin, Y., Du, E., Wei, Q., Li, Y., Huang, D., Wang, G., Veenstra, J.A., Li, S., Li, N., 2021. Genomics and peptidomics-based discovery of conserved and novel neuropeptides in the American Cockroach.

Journal of Proteome Research. https://doi.org/10.1021/acs.jproteome.0c00596

Supporting information S3: Gene structures of alternatively spliced *capa, itp, mip, ms, neuroparsin* and *ok* genes of *S. gregaria*. For the *ms* gene, a graphical representation of the gene structure (see Gremme et al., 2013) with resulting splice variants is given (A, B).



Detailed gene structures of alternatively spliced transcripts; including overall gene lenght and exon positions.

Gene	start	end
capa gene	1	114558
mRNA transcript a	1	114558
exon	1	86
exon	36329	36418
exon	65619	65770
exon	97016	97389
exon	114430	114558
mRNA transcript a	64	97257
exon	64	86
exon	36329	36417
exon	97016	97257
ion transport peptide gene	1	1906
mRNA transcript a	1	1229
exon	1	285
exon	1109	1229
mRNA transcript b	1	1906
exon	1	285
exon	1797	1906
myoinhibitory peptide gene	1	175315
mRNA transcript a	1	175315
exon	1	243
exon	388	426
exon	110015	110305
exon	175196	175315
mRNA transcript b	1	175315
exon	1	341
exon	110039	110305

Gene	start	end
neuroparsin gene	1	35032
mRNA transcript a	1	35032
exon	1	62
exon	2892	3007
exon	34887	35032
mRNA transcript b	1	29038
exon	1	62
exon	2892	3007
exon	28905	29038
mRNA transcript c	1	5003
exon	1	62
exon	2892	3007
exon	4870	5003
mRNA transcript d	1	7680
exon	1	62
exon	2892	3007
exon	7544	7680
orcokinin gene	1	369642
mRNA transcript a	1	369642
exon	1	75
exon	227733	227816
exon	257447	257469
exon	284180	284282
exon	369424	369642
mRNA transcript b	1	258143
exon	1	75
exon	227733	227816

exon	175196	175315
myosuppressin gene	1	4322
mRNA transcript a	1	2495
exon	1	225
exon	2398	2495
mRNA transcript b	1	4322
exon	1	225
exon	4234	4322

exon	256832	257541
exon	257938	258143

Reference:

Gremme, G., Steinbiss, S., Kurtz, S., 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Transactions on Computational Biology and Bioinformatics 10, 645–656.