Published in final edited form as:

Peptides. 2007 June ; 28(6): 1282–1291.

Neuropeptide Precursors in *Tribolium castaneum*

Andinet Amare and Jonathan V. Sweedler*

Department of Chemistry and the Institute of Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Abstract

Neuropeptides and neurohormones are among the more diverse and functionally important classes of cell-to-cell signaling molecules involved in animal development and behavior. Less is known about the hormones and neuropeptides of the red flour beetle, *Tribolium castaneum*, than many other insects. However, the genomic information becoming available from this organism presents an opportunity to identify multiple neuropeptide and hormone genes, and hence their associated protein precursors. Using similarity-based prediction, we report new neuropeptides and hormone precursors from *T. castaneum*, bringing the number of annotated precursors to 37. We identified one prohormone (SVDPIDGDLIG-containing) having little similarity to other insect prohormones. The conversion of the protein precursors into bioactive peptides requires a suite of processing enzymes and a number of enzymatic steps; using the web-based NeuroPred application and similarity-based bioinformatics approaches, we predict 132 likely peptides that may result from the enzymatic processing of these gene products.

Keywords

neuropeptides; NeuroPred; prohormone; *Tribolium*

1. Introduction

The red flour beetle, *Tribolium castaneum*, is among the species whose genome is being sequenced. Rationales for sequencing this genome include the insect's short generation time and demonstrated ease of genetic manipulation, the potential information to be gained about evolutionary events responsible for semi- and complete metamorphosis, and the opportunity to draw connections between non-human and human genomes, especially by filling the gaps created by genes that are lost in *Drosophila* [7]. The availability of the *Tribolium* genome also presents an opportunity to identify neuropeptide and hormone precursors from this organism.

Neuropeptides and hormones are signaling molecules that affect a wide range of behaviors and physiological functions; for example, they modulate the molting process in arthropods, essential for the growth and maturation of the organism [15]. Even after decades of research elucidating neuropeptide sequences from an organism, additional peptides often continue to be discovered [9,19,24]. The increasing availability of genomic information combined with sensitive analytical methods is not only greatly enhancing identification of neuropeptides in less-studied organisms, but also in well-studied model organisms. This advantage was clearly

*To whom correspondence should be addressed: Jonathan V. Sweedler, 600 S. Mathews Ave. 63-5, University of Illinois, Urbana IL 61801, jsweedle@uiuc.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

demonstrated in the honey bee, *Apis mellifera*, where the number of known prohormones was increased from three to 33 in a genome-assisted neuropeptide discovery study that employed mass spectrometry (MS) and bioinformatics tools [19]. Similar numbers of precursors were also identified in the fruit fly, *Drosophila melanogaster* [16,26], and African malaria mosquito, *Anopheles gambiae* [28], upon the availability of genomic data.

The identification of neuropeptide genes is one step towards understanding the functional connectivity of neuropeptides. However, complex processing events accompanying neuropeptide production involve proteolysis at select basic residues and non-conventional proteolytic sites, and numerous posttranslational modifications. These events can complicate the identification of neuropeptides, particularly when the method of identification relies on mass. Duckert et al. [12] have produced neural network-based cleavage prediction while our group [2,18,19,32] has developed statistics-based cleavage prediction algorithms at mono and dibasic residues that include several of the more prevalent posttranslational modifications. The prediction tools we have developed are accessible via a web-based application [31].

Interestingly, our recent data suggest that there are significant differences in the processing of monobasic and dibasic sites between animals from different phyla, likely because the specificity of the processing enzymes is different. Thus, we have developed phyla-specific models for insects (*Drosophila* and *Apis*) [30], mollusks (*Aplysia*) [18], and mammals [2]. The *Tribolium* genome represents an excellent opportunity to apply our insect model, trained with *Drosophila* and *Apis* prohormones, to predict the *Tribolium* neuropeptides processed from these precursors. As biochemical information becomes available, these correct (and incorrect) predictions will be used to further refine our insect neuropeptide processing model, allowing more accurate predictions with additional insects that will be sequenced in the future.

What is the significance of identifying prohormones from various species? The absence or presence of specific prohormones in a species can point to factors that may be responsible for unique aspects of organisms, such as what makes honey bees social and specific wasps solitary. Consider for example, examination of genes that are well characterized in one species, but absent in another, may lead to an understanding of compensatory mechanisms; or discovery of a precursor in a particular species may help to identify related precursors in others. Also, the understanding of unique genes in insects such as *Tribolium* may enable design of targeted pesticides.

Our objective is to identify neuropeptide and prohormone precursors from the *Tribolium* genome and apply our insect cleavage predictor to identify the most likely set of neuropeptides present in this organism; our predictions can be validated using direct biochemical assays in future studies. We have searched the *Tribolium* genome for precursors known in other insects, identifying 37 precursors and 132 putative peptides resulting from proteolysis at basic residues. One of these precursors has little similarity in other insects. Some expected precursors were not found in the genome, either because they were not sequenced and assembled yet, or because they are indeed missing from *Tribolium*.

2. Experimental Section

2.1. Data

The 2nd assembly of the *Tribolium* genome, sequenced to seven-fold redundancy, the expressed sequence tags (EST) and unassembled sequence library were obtained from the Human Genome Sequencing Center at the Baylor College of Medicine (BCM) ftp server (<http://www.hgsc.bcm.tmc.edu/projects/tribolium/>). The genomic library and 9 400 proteins generated from GLEAN automatic prediction [13] (referred to as Glean proteins) were also available from BCM to be queried using a web interface.

2.2 Prohormone profile

A profile was created for each prohormone based on multiple sequence alignments using the CLUSTALW [34] program available through the Biology Workbench version 3.2 (<http://workbench.sdsc.edu/index.html>). This profile was created in order to identify the conserved motifs of the prohormone and is used to evaluate the validity of predicted prohormones.

2.3 Similarity search

Similarity searches were conducted using the Basic Local Alignment Search Tool (BLAST) [1] either from the BCM page or from standalone BLAST. The settings for BCM BLAST are predetermined and unchangeable. The parameters for standalone BLAST were modified depending on the length of the query. For a long query, only the expectation value (e-value) was increased. Occasionally the filtering option, which masks low-complexity sequences marked by biased representation of certain residues (including, acidic, basic or proline stretches) in the query sequence, was turned off. For short queries (e.g., only when the query was a short peptide), the e-value was raised to 20 000, the word size was decreased to 2, gap penalty was reduced to 9, and the filter was turned off. The BLASTp program was selected when searching for a similar protein to the query protein in a protein database (e.g., GLEAN protein collections); likewise, tBLASTp was used when searching against a nucleotide sequence with a protein query. tBLASTn is particularly useful when the protein of interest is not found in the protein database. By selecting tBLASTn, we ensure sensitive search against the nucleotide database because nucleotide sequences are first translated into protein sequences prior to sequence comparison.

2.4 Evaluation of BLAST hits

BLAST hits with the lowest e-value were examined first. If the match between query and hit included the conserved motif of the query (see Section 2.2), the hit was tagged for further analysis. If the query was performed against a protein database, then the protein sequence was copied from the BMC page. If a nucleotide database was queried, a total length of 10 000 base pairs surrounding the hit was extracted using PERL scripts. This sequence was used with the gene prediction programs (Section 2.5).

2.5 Gene prediction

The FGENESH (www.softberry.com) and Augustus (<http://augustus.gobics.de/submission>) gene prediction programs were used to predict genes for potential BLAST hits. These programs were selected because they have algorithms trained for *Tribolium*. Both programs provide the capacity to output alternate predictions.

2.6 Evaluation of predicted genes

The following two approaches were used to evaluate the validity of the predicted proteins as homologs of the query sequence.

2.6.1 Profile-profile alignment—Whenever a prohormone profile was available for the query sequence, a sequence-to-profile alignment (CLUSTALWPROF from Biology Workbench) was performed with the potential protein against the profile. A good alignment against the profile preserves the consensus for the prohormone.

2.6.2 Position Specific Iterative–BLAST (PSI-BLAST)—A second way to determine if a proposed protein belongs to the prohormone family is to use that protein as a query against the database. If the reported top hits belong to the prohormone family, one can conclude that the query also belongs to the same family. For this comparison, the PSI-BLAST program with five iterations and default Biology Workbench settings were used, often against the GenBank

invertebrate database but occasionally against the SwissProt database. A PSI-BLAST result relies on the initial BLAST matches to the query it identified from the search database. Because a BLAST score is dependant on the length of the query sequence and the size of the database, PSI-BLAST search was often conducted in the GenBank invertebrate database where evolutionary closeness and limitedness of the database is expected to yield the best hits at the first iteration, thereby ensuring good matches for subsequent iterations. The SwissProt database, containing proteins from a broad range of phyla, was queried when the search in the GenBank invertebrate database was unsuccessful in that alignments to the query sequence were insignificant.

2.7 Prohormone features

Once a reasonable prohormone was predicted, SignalP version 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) [5] was used to identify the signal peptide, which every prohormone needs to have in order to proceed through the regulated secretory pathway [8]. Prohormones often carry basic residues that are frequently cleaved to generate smaller peptides. The exact cleavage sites were predicted using a cleavage prediction method trained on *Drosophila* and *Apis* prohormones [32]. The parameters used for cleavage prediction were: Intercept (-11.373); P1 (Lys = -1.4129); P2 (Gly = 5.7359; Lys = 12.2048; Arg = 9.3092); P4 (Arg = 2.1872); P6 (Phe = 3.4665); P'1 (Ser = 6.267); P'2 (Pro = 2.5695); P'3 (Lys = 2.1654) and P'4 (Leu = 3.3512; Ser = 3.3132).

3. Results

We identified 34 prohormones and three defensin precursors from *Tribolium* based on similarity searches from known insect precursors (Table 1 and Supplemental Information). Thirty-three of these precursors existed among the Glean proteins available through the BCM Sequencing Center, although only eclosion hormone-2 (EH-2), bursicon alpha and bursicon beta were assigned descriptions referring to their function while the rest were labeled as unknown proteins. Functionally annotated versions of sulfakinin and eclosion hormone-1 (EH-1) prohormones were annotated in the NCBI database. Three prohormones, CRF-related diuretic hormone-II (DH-II CRF related), neurophysin and short neuropeptide F (BK006117), were newly identified and annotated in this study. Of the precursors predicted, we discuss several in detail below because they are novel neuropeptides, highlight issues in the annotation process or present interesting aspects of neuropeptides.

3.1.1 SVDPIDGDLIG-containing—A precursor from *Apis*, showing weak similarity with orckinin from the red swamp crayfish, *Procambarus clarkia*, [19] was queried against the *Tribolium* genome. A Glean protein (Glean 05944) was found to match this precursor (e-value $1e-7$). Glean 05944, a 291 amino acid precursor, contains a signal peptide of 19 residues and several basic residues. Assuming all cleavages predicted by NeuroPred are correct, and subsequent processing events including amidation take place, this protein potentially yields five copies of Ser-Val-Asp-Pro-Ile-Asp-Gly-Asp-Leu-Ile-amide, two copies of Ser-Leu-Asp-Arg-Ile-Gly-Gly-Gly-Asn-Leu-Val-amide, and single copies of Lys-Leu-Ser-Cys-Ala-Thr-Leu-His-Ile-Leu-Gly-Arg-Gln-Trp-Ser-Arg-Leu-Phe-amide, Ser-Val-Asp-Pro-Ile-Asp-Gly-Asp-Asp-Leu-Ile-amide, and Ser-Leu-Asp-Gly-Ile-Gly-Gly-Gly-Asn-Leu-Val-Gly-Arg-Gly-Val-Asp-Pro-Ile-Asp-Gly-Asp-Leu-Ile-amide, as well as several non-amidated peptides. Although this precursor was identified using an *Apis* precursor, it is significantly different from the query in prohormone structure as well as sequence (Figure 1). The *Apis* precursor lacks the repetitive sequences that are found in *Tribolium*. A PSI-BLAST search against the GenBank invertebrate database identified LWamide as a possible homolog, but closer examination revealed that the high score was due to repetitive occurrences of Gly-Lys-Arg motifs in both sequences. Hence, it appears that Glean 05944 may be a unique protein to *Tribolium*.

3.1.2 Neurophysin—A BLAST search against the *Tribolium* genome using known rat and mouse prohormone precursors indicated that *Tribolium* contains a precursor that shows a strong similarity to rat and mouse neurophysin-1 and neurophysin-2. This *Tribolium* precursor gene showed strong similarity to vertebrate neurophysins as well as molluscan neurophysin equivalents (cephalotocin and Lys-conopressin). Adjacent to the signal peptide lies the peptide Cys-Leu-Ile-Thr-Asn-Cys-Pro-Arg-Gly-Gly, which was isolated from the migratory locust, *Locusta migratoria*, and labeled as an Arg-vasopressin-like peptide [14,27]. Because both neurophysin-1 and neurophysin-2 from mouse and rat were used as queries against the *Tribolium* genome, and both identified the same nucleotide stretch on contig CM000283, there may be only one neurophysin present in *Tribolium*. Neurophysin has not been found in *A. mellifera* and *D. melanogaster*. Although neurophysin has been indicated in *L. migratoria*, this is the first report of a full-length neurophysin precursor from insects.

3.1.3 Prothoracicotropic hormone (PTTH)—PTTH controls the release of ecdysteroid hormone from thoracic glands at molting [22,23]; it can also end diapausing in insects [38]. PTTH is a homodimeric molecule where each chain contains seven Cys involved in inter- and intra-disulfide bond formation [20]. Similarity searches using various insect PTTH precursors identified two potential precursors. One of the precursors from CM000276.1 (Glean 04893) showed the greatest similarity with a partial mRNA sequence of the African malaria mosquito, *A. gambiae*, ENSANGG00000021732 (gi: 57968811, ref: XM_563293, e-value 2e-45) followed by a *Drosophila* trunk protein (gi:19550176; ref:NM_057419.2, CG5619-RA, e-value 6e-36). It also weakly matched various PTTHs, especially from corn earworm, *Helicoverpa zea*, and cotton bollworm, *Helicoverpa armigera*. Another precursor from CM000277.1 lacking a signal peptide was also identified, matching ENSANGP00000027261 from *A. gambiae*, (e-value 8 e-10), preproPTTH from beet armyworm, *Spodoptera exigua*, and domestic silkworm, *Bombyx mori*, (e-value 2e-7), as well as PTTH from other insects. A short, 64-residues-long protein (Glean 00735) matching the first 37 residues of a FGENESH-predicted protein was also identified.

3.1.4 Neuropeptide F (NPF)—Two forms of NPF have been annotated, long NPF and short NPF. Short NPFs with one copy of Pro-Xxx-Leu-Arg-Leu-Arg-Phe-amide, Pro-Xxx-Leu-Lys-Thr-Arg-Phe-amide or Arg-Phe-Arg-Phe-amide, where Xxx is any amino acid sequence in the C-terminal, were found in the yellow fever mosquito, *Aedes aegypti* (three peptides) [25,36, 37], Colorado potato beetle, *Leptinotarsa decemlineata* (two peptides) [33], *H. zea* (two peptides) [17], locust, *Schistocerca gregaria* (one peptide) [11,29], and *D. melanogaster* (four peptides) [16,35]. In *Tribolium*, a precursor encoding a Ser-Pro-Ser-Leu-Arg-Leu-Arg-Phe-amide, which is identical to MS confirmed *Apis* sNPF peptide [19] has been annotated for the first time.

3.1.5 Eclosion hormone (EH)—Two EH precursors were predicted from *Tribolium* and located on different genes, both displaying similarity to EH from Asian corn borer, *Ostrinia furnacalis*, to tobacco hornworm, *Manduca sexta*, and, *H. armigera*. Glean 00178 was similar to one of the EHs (from CM000277.1), but without the C-terminal exon, lacked three Cys, which are necessary for proper folding [21]. The FGENESH prediction truncated at the N-terminal and therefore, lacked a signal peptide. To reconcile this difference, the missing N-terminal segment was appended to it. We searched the GenBank and SwissProt databases for reported cases of two or more EHs per species. Three EHs were predicted from *A. aegypti*, of which two (Q16QN2_AEDAE and Q16QN3_AEDAE) are located on contig CH47741, and the other (Q16GE3_AEDAE) is located on contig CH478286. Apart from *A. aegypti*, *Tribolium* is the only other insect for which multiple EH genes are reported.

3.2 Precursors lacking signal peptides

Proteins that follow the secretory pathway carry a signal peptide that directs them to the endoplasmic reticulum to continue translation and later be transported to the Golgi apparatus where they will be modified, sorted and packaged. Thus, the signal peptide is a natural feature of a prohormone precursor, and its absence from prediction could be attributed to a gene predictor's choice of a start site. Such a discrepancy can be resolved by using a different gene prediction model or by considering other potential start sites. EH (Section 3.1.5) is an example where FGENESH and GLEAN have been used complementarily to assign a signal peptide sequence to the prohormone. If the genome is not sufficiently extended at the 5' region (perhaps due to an incomplete sequence), the proper start site may not be identified, resulting in a partial prohormone sequence lacking the signal sequence. Three prohormones predicted as Glean and FGENESH proteins (calcitonin-related diuretic hormone DH-31 (Glean 04987), PTTH (Glean 00735), neuropeptide protein-like-1 (Glean 06787)) and DH-II corticotropin releasing factor-related (only FGENESH predicted) lack an identifiable signal peptide, even though similarity searches and sequence profiles identified the precursors. A possible explanation for the lack of signal peptide is that the DNA sequence encoding for the signal peptide is not available, i.e., is a part of the DNA sequence marked by series of N's. However, for all three prohormones, the start of the transcription site and the unsequenced stretch upstream of the start site is separated by more than 4 000 base pairs. Given that the largest intron size observed was 2 299 base pairs for the *Tribolium* ITGQGNRIF-containing prohormone (inferred from data in Table 1), limiting the search for unsequenced regions to 4 000 base pairs upstream of the predicted start site is reasonable. If the signal peptide is missing because relevant DNA sequences were not assembled correctly, this error may be rectified in the future. If the lack of signal peptide is due to the annotation process, in which transcription start sites were not predicted correctly, these issues will be resolved by using different algorithms for finding transcription start sites. Lastly, the possibility remains that both the sequencing and annotation are correct; in this case, the final protein lacks a signal peptide and thus is not a neuropeptide prohormone.

3.3 Precursors not found in *Tribolium*

Using similarity-based analysis, neuropeptides corresponding to amnesiac peptide, apidaecin, corazonin, hugin, pigment dispersing hormone, neurokinin, long NPF, neuropeptide FF, homologues to *Apis* neuropeptide protein-2 and -3, and RFamide-1 and -2 were not found. Motif and BLAST searches for short but conserved sequences were employed for some peptides such as baratin, corazonin, and pigment-dispersing hormone, but these searches were not successful. Other short peptides without complete precursor sequences discovered from other species, such as ecdysiotropin, trypsin-modulating oostatic factor, and cAMP-generating peptide of the flesh fly, *Sarcophaga bullata*, were also not found in this study.

Long NPFs share the same C-terminal Gly-Arg-Pro-Arg-Phe-amide sequence and are 36–39 residues long. Distinct genes for precursors of the short and long NPF have been identified in the fruit fly, *D. melanogaster* [6,16,35]. Long NPF is considered a member of the neuropeptide Y (NPY) family because of the sequence similarity existing between NPF and NPY precursors, and between their respective receptors. Similarity searches against the *Tribolium* genome, EST library and unassembled sequences using invertebrate NPF and vertebrate NPY did not reveal NPF homologs.

There are a number of reasons why prohormone precursors may be missing from *Tribolium*. One possibility is that the prohormone gene is located within the portion of the gene that is not sequenced yet. Currently 152 Mb of the estimated 158 Mb size of the *Tribolium* genome has been sequenced [10]. The receptor for *Tribolium* pigment dispersing hormone (XM_966645.1) is identified even when the ligand is missing, suggesting the possibility that the prohormone gene is not yet sequenced. On the other hand, the corazonin prohormone as well as its receptor

are missing from the *Tribolium* genome, whereas either corazonin precursors or its receptors are found in insects from Diptera, Lepidoptera, and Hymenoptera (e.g., prohormones are identified from the honey bee, *A. mellifera* (BAD90662), the fruit fly, *D. melanogaster* (Q26377), the African malaria mosquito, *A. gambiae* (XP_001230939.1), *B. mori* (BAC66443.1) and the greater wax moth, *Galleria mellonella* (Q9GSA4) and receptors from *D. melanogaster* (AAN100045), *A. aegypti* (EAT36445.1), and *M. sexta* (AAR14318.1)), suggesting that perhaps corazonin is missing from the Coleoptera to which *Tribolium* belong. Finally, small peptides whose precursors are not known may simply be too difficult to identify by homology searches or unique only to the organism from which they were originally isolated. As more experimental data become available, and the sequencing project is completed and gaps in the genomic information filled, we feel that several of these missing precursors will either be discovered, or their absence in the *Tribolium* genome confirmed.

3.4 Comparison of FGENESH- and GLEAN-predicted prohormones

Every prohormone reported in this paper was predicted using FGENESH. For many of the prohormones, the GLEAN and FGENESH predictions were identical. However, some differences were observed, including varying start sites that resulted in variations in signal peptide length and the number and length of predicted exons. The choice of a start site in FGENESH tended to result in short signal peptides. In the case of FGENESH-predicted adipokinetic hormone (AKH)-I and EH (equivalent to Glean 00178), the signal peptides were too short to be recognized as such by SignalP. In other cases, the signal peptide may be too long, thereby producing the reverse effect. For example, Glean 13258 was identical to FGENESH neuroparsin prohormone except that it was extended at the N-terminal side by 29 residues, which was sufficient for it to be predicted as lacking signal peptide, while the FGENESH equivalent was predicted otherwise.

The variations in exon prediction were sometimes easy to identify when the predicted precursor was too long compared to the other known precursors of that family. Several such cases are illustrated here. The GLEAN predictions for AKH-I (Glean 16051), AKH-II (Glean 11561) and ion transport protein (Glean 0429) were long, and were the result of artifactual merging with other proteins. The PTTH precursor (Glean 00735) was only 64-residues-long and was identical to the FGENESH PTTH for 37 residues at the N-terminal. In a separate case, the EH (Glean 00178) was missing the C-terminal exon, thus lacking three Cys that are necessary for proper folding. A precursor showing strong similarity to the pheromone biosynthesis activating neuropeptide/diapause hormone family was identified (CM000276.1, Glean 04636). The prohormone predicted using FGENESH is C-terminally extended by 66 residues. The FGENESH algorithm predicted a precursor from three exons, whereas GLEAN predicted only two exons. The first exon, which is the same in both predictions, encodes a 27-residues-long peptide (of which the first 24 residues belong to the signal peptide). The GLEAN prediction predicts an extra eight residues at the end of the second exon, which do not appear on the FGENESH prediction. FGENESH predicted a third exon containing residues 90–162, and encodes three pyrokinin-like peptide motifs, each followed by a monobasic Arg (His-Val-Val-Asn-Phe-Thr-Pro-Arg-Leu-Gly; Glu-Ser-Gly-Glu-Glu-Phe-Val-Asn-Asn-Ala-Pro-Glu-Asp-Arg-Trp-Leu-Gln-Asn-His-Glu-Thr-Ser-Gly-Glu-Met-Leu-Tyr-Gln-Arg-Ser-Pro-Pro-Phe-Ala-Pro-Arg-Leu-Gly; His-Ser-Ser-Pro-Phe-Ser-Pro-Arg-Leu-Gly). The first and second exons together encode for another peptide followed by Arg-Lys-Lys-Arg (Thr-Pro-His-Glu-Ser-Ser-Val-Pro-Asn-Glu-Arg-Asn-Asp-Asp-Ser-Lys-Glu-Thr-Tyr-Phe-Trp-Phe-Gly-Pro-Arg-Leu-Gly).

3.5 Statistics-based peptide prediction

Using the statistics-based prohormone processing at basic residues, we predicted 132 peptides from the 37 prohormones. Some of the peptides such as AKH-I to AKH-III (Gln-Leu-Asn-

Phe-Ser-Thr-Asp-Trp-Gln; Gln-Val-Thr-Phe-Ser-Arg-Asp-Trp-Asn-Pro-Gly and Gln-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly, respectively), various peptides of allatostatin B (Trp-Asn-Lys-Asp-Leu-His-Ile-Trp-Gly; Gly-Trp-Asn-Asn-Leu-His-Glu-Gly-Trp-Gly; Ala-Trp-Gln-Ser-Leu-Gln-Ser-Gly-Trp-Gly; Asn-Trp-Gly-Gln-Phe-Gly-Gly-Trp-Gly; Ser-Lys-Trp-Asp-Asn-Phe-Arg-Gly-Ser-Trp-Gly; Glu-Pro-Ala-Trp-Ser-Asn-Leu-Lys-Gly-Ile-Trp-Gly) and sulfakinin (Gln-Thr-Ser-Asp-Asp-Tyr-Gly-His-Leu-Arg-Phe-Gly; Gly-Glu-Glu-Pro-Phe-Asp-Asp-Tyr-Gly-His-Met-Arg-Phe-Gly) are surrounded by dibasic sites known to be cleaved at a high frequency and result in peptides showing strong similarity to other known peptides in these families. Hence, there is confidence in these proteolytic predictions. Often the challenge in prediction of cleavage is at single Lys or Arg sites. Several instances of putative single cleavage sites are observed in *Tribolium*. For example, FLRFamide prohormone, which contains four copies of Phe-Leu-Arg-Phe-Gly and one copy of Phe-Ile-Arg-Phe-Gly, each followed by a single Arg, were all predicted as cleaved, whereas pyrokinin-like precursor, which carries three Pro-Arg-Leu-Gly-Arg sequences, was not predicted as cleaved. The triad Pro-Arg-Leu-amide is functionally important in periviscerokinins and ecdysis triggering hormone [26], and one would expect that peptides in pyrokinin containing the motif Pro-Arg-Leu-Gly are also functional. Thus, the peptides predicted with the statistical algorithm can be taken as initial points of analyses, where most cleavages are expected to be confirmed as true cleavages and some will undoubtedly be incorrect.

4. Discussion

The annotation of 37 neuropeptide and hormone precursors and the prediction of more than 100 neuropeptides advance our information on *Tribolium* neuropeptides. Again, the utility of similarity-based neuropeptide gene identification, previously illustrated in *Drosophila* [3,4, 24], *Apis* [19], and *Anopheles* [28], is reiterated here in *Tribolium*. Because BLAST scoring schemes depend on the length of the query sequence and the database referenced, short neuropeptides whose precursors have poor similarity with other insects often score low (have high e-values). Limiting the search space increases the chance of locating biologically meaningful matches. The search for orcokinin precursor from *A mellifera* (Section 3.1.1), demonstrates that a high score does not necessarily correspond to the correct match, yet close examination of the match often helps in identifying a novel precursor.

Similarity searches identify proteins related to the query sequence and thus, have limited capacity to find prohormones that are unique to a particular organism. In such cases, biochemical characterization is essential. For example, complementary neuropeptide motif searches and MS were employed in the annotation of the *Apis* neuropeptide precursors [19] and enabled the discovery of 10 precursors with no similarity to known proteins. Recently, Liu and coworkers [24] also employed specific neuropeptide features such as the occurrence of Gly-Lys-Arg to mine the *Drosophila* proteins, and identified 76 unannotated putative peptide genes, several of which they confirmed by MS.

While multiple neuropeptides are predicted, additional neuropeptides and their precursors will undoubtedly be discovered in the future, as will differences between our statistical-based predictions and experimentally confirmed peptides (when such information becomes available). Neuropeptide discovery is most effective when the *in silico* methods employed here are complemented with biochemical measurements.

Here two gene prediction methods, FGENESH and Augustus, have been used and compared to the protein predictions with GLEAN. Augustus, which uses a hidden Markov model and generates multiple splice variants, and was trained on known or verified *Tribolium* genes, was found to miss several genes that were identified by FGENESH and GLEAN. FGENESH, trained on verified genes from *Tribolium* and other related organisms, was found to predict

short genes and many that were missed by Augustus. GLEAN, first employed in the honey bee genome project, was created to consolidate and report a consensus model by comparing results from several gene prediction algorithms [13]. The GLEAN proteins reported are the unified results of Augustus, FGENESH, NCBI, and Ensemble gene predictions. Elisk et al. [13] reported that GLEAN's predictions were slightly better than individual prediction tools, as does the *Tribolium* sequencing consortium. Likewise, others have reported that GLEAN predictions are slightly better than individual predictions [10]. This raises the question of whether our predictions, which use a single prediction tool, are more accurate than GLEAN. Although we use primarily FGENESH as the gene prediction model, both the initial query into the *Tribolium* database and gene verification rely on human knowledge and judgment, which may improve performance. In fact, this is why computer-predicted genes are normally verified by manual annotation. We have identified cases where the GLEAN prediction, viewed from the existing knowledge of prohormone structure, are questionable, and have also indicated cases where GLEAN and FGENESH predictions are identical. The validity of the various predictions for the neuropeptides prohormones awaits experimental verification.

6. Supplemental Information

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This material is based upon work supported by the National Institute on Drug Abuse under Award No. P30 DA 018310 to the UIUC Neuroproteomics Center, and the National Institutes of Health under Award No. R01 NS31609.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10. [PubMed: 2231712]
2. Amare A, Hummon AB, Southey BR, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV. Bridging neuropeptidomics and genomics with bioinformatics: Prediction of mammalian neuropeptide prohormone processing. *J Proteome Res* 2006;5:1162–67. [PubMed: 16674105]
3. Baggerman G, Boonen K, Verleyen P, De Loof A, Schoofs L. Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J Mass Spectrom* 2005;40:250–60. [PubMed: 15706625]
4. Baggerman G, Liu F, Wets G, Schoofs L. Bioinformatic analysis of peptide precursor proteins. *Ann N Y Acad Sci* 2005;1040:59–65. [PubMed: 15891006]
5. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95. [PubMed: 15223320]
6. Brown MR, Crim JW, Arata RC, Cai HN, Chun C, Shen P. Identification of a *Drosophila* brain-gut peptide related to the neuropeptide Y family. *Peptides* 1999;20:1035–42. [PubMed: 10499420]
7. Brown, S.; Denell, R.; Beeman, R.; Richard Gibbs, R. Rationale to sequence the genome of the red flour beetle, *Tribolium castaneum*. <http://www.hgsc.bcm.tmc.edu/projects/Tribolium/TriboliumWhitePaper.pdf>
8. Campbell, N. *Biology*. 4. Menlo Park: The Benjamin/Cummings publishing company, Inc; 1996.
9. Clynen E, Huybrechts J, Verleyen P, De Loof A, Schoofs L. Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry. *BMC Genomics* 2006;7:201. [PubMed: 16899111]
10. Consortium TGS. The first genome sequence of a beetle, *Tribolium castaneum*, a model for insect development and pest biology. Submitted
11. De Loof A, Baggerman G, Breuer M, Claeys I, Cerstiaens A, Clynen E, et al. Gonadotropins in insects: an overview. *Arch Insect Biochem Physiol* 2001;47:129–38. [PubMed: 11418931]

12. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel* 2004;17:107–12. [PubMed: 14985543]
13. Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM. Creating a honey bee consensus gene set. *Genome Biol* 2007;8:R13. [PubMed: 17241472]
14. Friedel T, Loughton BG, Andrew RD. A neurosecretory protein from *Locusta migratoria*. *Gen Comp Endocrinol* 1980;41:487–98. [PubMed: 6773849]
15. Hesterlee S, Morton DB. Insect physiology: the emerging story of ecdysis. *Curr Biol* 1996;6:648–50. [PubMed: 8793284]
16. Hewes RS, Taghert PH. Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. *Genome Res* 2001;11:1126–42. [PubMed: 11381038]
17. Huang Y, Brown MR, Lee TD, Crim JW. RF-amide peptides isolated from the midgut of the corn earworm, *Helicoverpa zea*, resemble pancreatic polypeptide. *Insect Biochem Mol Biol* 1998;28:345–56. [PubMed: 9692236]
18. Hummon AB, Hummon NP, Corbin RW, Li L, Vilim FS, Weiss KR, et al. From precursor to final peptides: a statistical sequence-based approach to predicting prohormone processing. *J Proteome Res* 2003;2:650–56. [PubMed: 14692459]
19. Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* 2006;314:647–49. [PubMed: 17068263]
20. Ishibashi J, Kataoka H, Isogai A, Kawakami A, Saegusa H, Yagi Y, et al. Assignment of disulfide bond location in prothoracicotropic hormone of the silkworm, *Bombyx mori*: a homodimeric peptide. *Biochemistry* 1994;33:5912–19. [PubMed: 8180220]
21. Kataoka H, Li JP, Lui AS, Kramer SJ, Schooley DA. Complete structure of eclosion hormone of *Manduca sexta*. Assignment of disulfide bond location. *Int J Pept Protein Res* 1992;39:29–35. [PubMed: 1634328]
22. Kopec S. Studies on the necessity of the brain for the inception of insect metamorphosis. *Biol Bull Woods Hole* 1992;42:322–42.
23. Kopec S. Studies on the necessity of the brain for the inception of insect metamorphosis. *Biol Bull Woods Hole* 1922;42:322–42.
24. Liu F, Baggerman G, D'Hertog W, Verleyen P, Schoofs L, Wets G. *In silico* identification of new secretory peptide genes in *Drosophila melanogaster*. *Mol Cell Proteomics* 2006;5:510–22. [PubMed: 16291998]
25. Matsumoto S, Brown MR, Crim JW, Vigna SR, Lea AO. Isolation and primary structure of neuropeptides from the mosquito, *Aedes aegypti*, immunoreactive to FMRFamide antiserum. *Insect Biochem* 1989;19:277–83.
26. Nassel DR. Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog Neurobiol* 2002;68:1–84. [PubMed: 12427481]
27. Proux JP, Miller CA, Li JP, Carney RL, Girardie A, Delaage M, et al. Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria*. *Biochem Biophys Res Commun* 1987;149:180–6. [PubMed: 3689410]
28. Riehle MA, Garczynski SF, Crim JW, Hill CA, Brown MR. Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* 2002;298:172–5. [PubMed: 12364794]
29. Schoofs L, Clynen E, Cerstiaens A, Baggerman G, Wei Z, Vercammen T, et al. Newly discovered functions for some myotropic neuropeptides in locusts. *Peptides* 2001;22:219–27. [PubMed: 11179815]
30. Southey B, Hummon A, Richmond T, Sweedler J, Rodriguez-Zas S. Prediction of neuropeptide cleavage sites in insects. 2007submitted
31. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. *Nucleic Acids Res* 2006;34:W267–72. [PubMed: 16845008]
32. Southey BR, Rodriguez-Zas SL, Sweedler JV. Prediction of neuropeptide prohormone cleavages with application to RFamides. *Peptides* 2006;27:1087–98. [PubMed: 16494967]

33. Spittaels K, Verhaert P, Shaw C, Johnston RN, Devreese B, Van Beeumen J, et al. Insect neuropeptide F (NPF)-related peptides: isolation from Colorado potato beetle (*Leptinotarsa decemlineata*) brain. *Insect Biochem Mol Biol* 1996;26:375–82. [PubMed: 8814784]
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–80. [PubMed: 7984417]
35. Vanden Broeck J. Neuropeptides and their precursors in the fruitfly, *Drosophila melanogaster*. *Peptides* 2001;22:241–54. [PubMed: 11179818]
36. Veenstra JA. Isolation and identification of three RFamide-immunoreactive peptides from the mosquito *Aedes aegypti*. *Peptides* 1999;20:31–8. [PubMed: 10098621]
37. Veenstra JA, Costes L. Isolation and identification of a peptide and its cDNA from the mosquito *Aedes aegypti* related to *Manduca sexta* allatotropin. *Peptides* 1999;20:1145–51. [PubMed: 10573285]
38. Wei ZJ, Zhang QR, Kang L, Xu WH, Denlinger DL. Molecular characterization and expression of prothoracicotropic hormone during development and pupal diapause in the cotton bollworm, *Helicoverpa armigera*. *J Insect Physiol* 2005;51:691–700. [PubMed: 15869761]

```
>7619888_7619889
```

MTAQMFTTIALLLSLSAIAAA GTIKTAPARTPSTQDDASFPDGPVKRFD AFTTGFGHSKR NFDEIDRSG
FGFAKK NFDEIDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDE
IDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDEIDRSGGFGFNKR NFDEIDRSGGFGV KRVYVPRYIANLY KRN
FDEIDRSGGFGFNKR NFDEIDRTGFGF HKRDYDVFPDKRN NFDEIDRSGGFGVRRNVE

>JVS_027_Apis_Orcokinin_GB11

MPRHSYFALSIALSITATVWVPIPTVQAEETNLLRREFYGPVNPELFAAFLLDDHAI I IRLRKSTLARKKGFV
GDGDERIFDLTILTGSKRRMDEFSGLPREKVGQSDRYLSRLRGGGNVARLPARSGVGKSNYDARDGQDER
LSRNLDDQIGAGNLLRGASSLSAEREGSSYDGRWSRRLNDQIGGGNLLREADSRGERNLDSIGGGNLVRGA
NLRLLDRNLDDQIGGGNLLRVS LTD PARRPVETVHARGRENQRDFSSGSGSTNELVDELSPVSERETLERFGR
NIDEIDRTAFDNFFKRNLD EIDRVGWSGFVKRLTNYLATTGHGTNTGGPVLTRRFG

```
>gi|74477180|gb|CM000283.1||[3770000-
```

3780000] | FGENESH_SVDPIDGDLIG_containing_peptide signalp=19

MRFVTVAIVIVFAAAAAEFAAPKPAGLSGYSRVLCNFSRKRKLSCATLHILGRQWSRLFGRSVDPIDGDLIGRSVDP
IDGDLIGRSVDPIDGDLIGRSLDRIGGGNLVGRSVDPIDGDLIGRSVDPIDGDDLIGRSLDRIGGGNLVGRSVDPID
GDLIGRSLDGIGGGNLVGRGVDPIDGDLIGRSLDIKRLDGYRRKHNAYEEVIGQKTIRNFGVLQLGGGYGVAKRFA
QPGKYDTKTEKHRRGPLNGLIPGGAFGRAARSCLKYFSFDSRQATKPGIYKTYQLLEME

Figure 1.

Comparison of prohormone organization and sequences for the red swamp crayfish, *Procambarus clarkii* orcokinin (7619888_7619889), the honey bee, *A. mellifera*, orcokinin and SVDPIDGDLIamide-containing precursors (this study). Signal peptide is indicated with dashed line, and solid underline indicates peptides predicted using a cleavage prediction model trained on *Apis* and *Drosophila* precursors. Double underline indicates peptides confirmed by MS analysis. Orcokinin contains 13 highly similar peptides and two additional peptides. The repetitive occurrence of short peptides is also displayed in the SVDPIDGDLIG-containing prohormone. In contrast, the orcokinin-related *Apis* prohormone contains five peptides, of which two show considerable sequence similarity.

Prohormones identified from *Tribolium castaneum*. The list of prohormones identified and annotated in this study is given, as well as the scaffold number and exon locations. The gene prediction algorithm used for the particular prohormone and the equivalent Glean protein is also indicated. The validity of the predicted prohormone is confirmed by using the prohormone as a query sequence for a PSI-BLAST search or from Profile-Profile alignment using CLUSTALPROF. Also, the closest homolog for the precursor determined by BLAST searches is indicated by accession number, organism name and e-value.

Table 1

Prohormones ^a	Scaffold Genbank accession number	ORF location	Glean	Prediction Algorithms			PSI- BLAST	CLUSTAL WPROF	Confirmation	BLAST	E- value	
				FGENESH	Augustus						Accession #	Organism
1	Adipokinetic hormone-1 (AKH-I)	9204-9374 9804-9842	16051	x	x		x	x	gi116117450 gb E AU75900.1	<i>A. gambiae</i>	5e-05	
2	Adipokinetic hormone-2 (AKH-II)	15471-15552 15643-15804 15853-15869	11561	x	x		NA	x	gi108872779 gb E AT37004.1	<i>A. aegypti</i>	0.52	
3	Adipokinetic hormone-3 (AKH-III)/ RPCH related	16684660-16684427 16683552-16682973	13125	x	NA		NA	x	NA	NA	NA	NA
4	Allatostatin B (ASTB)	1937490-1937368 1935093-1934680	07661	x	x		x	NA	gi108871490 gb E AT35715.1	<i>A. aegypti</i>	1e-18	
5	Allatostatin C (ASTC)	14499812-14499705 14499598-14499359	05428	x	x		x	NA	gi17981755 refl NP_523542.1	<i>D. melanogaster</i>	5e-06	
6	Allatotropin (AT)	9444629-94444841 9447644-9447738	06446	x	NA		x	x	gi167527245 gb AA Y68370.1	<i>S. cynthia</i>	4.4	
7	Bursicon (BSN) alpha	2892783 - 2892762 2892715-2892604 2892469 -2892202	13832	x	x		x	x	gi124648796 refl NP_650983.1	<i>D. melanogaster</i>	8e-39	
8	Bursicon (BSN) beta	2895286-2895175 2895125-2895000 2894314-2894145	13831	x	x		x	x	gi154645711 gb E AL34449.1	<i>D. pseudoobscura</i>	2e-50	
9	Calcitonin-like diuretic hormone (DH31) Diuretic hormone II	372082-372024 370903-370768	04987	x	NA		NA	NA	gi18928102 sp P82372 DIUX_DIPPU	<i>D. punctata</i>	1e-06	
10	Crustacean cardioactive peptide (CCAP)	5766594-5766747 5766962-5767239	01088	x	x		x	x	gi151701330 sp Q75UG5 CCAP_PERAM	<i>P. americana</i>	1e-20	

Prohormones ^a	Scaffold Genbank accession number	ORF location	Glean	Prediction Algorithms		PSI- BLAST	CLUSTAL WPROF	Confirmation		
				FCGENESH	Augustus			Accession #	Organism	E- value
11	Crustacean hyperglycemic hormone (Ion transport protein (ITP))	640642-640678	09429	x	NA	x	x	gi 6016403 sp Q26492 ITPL_SCHGR	<i>S. gregaria</i>	3e-35
		642764-642864								
		643568-643666								
		644265-644521								
		646173-646297								
		646566-646693								
12	Defensin-I	167939-167867	10517	x	NA	x	NA	gi 2493576 sp Q27023 DEF1_TENMO	<i>T. molitor</i>	1e-18
		167817-167641								
13	Defensin-II	6999518-6999611	12469	x	NA	x	NA	gi 13625795 gb AAK35160.1 AF311352.1	<i>A. luxuriosa</i>	2e-20
		6999666-6999823								
14	Defensin-III	4483461-4483690	06250	x	x	x	NA	gi 2493576 sp Q27023 DEF1_TENMO	<i>T. molitor</i>	3e-14
		4483740-4483908								
15	Diuretic hormone (DH-37)	68722-8988	02243	x	NA	x	NA	gi 3913490 sp P56618 DIUH1_TENMO	<i>T. molitor</i>	4e-07
		69477-69621								
16	Diuretic hormone-II DH-II CRF related	72070-72174	NA	x	NA	x	x	gi 3913491 sp P56619 DIUH2_TENMO	<i>T. molitor</i>	9e-17
		73227-73457								
17	Ecdysis triggering hormone (ETH)	1333176-1333382	09466	x	x	x	NA	gi 118778551 ref XP_308702.2 	<i>A. gambiae</i>	0.003
		1333435-1333474								
		1333523-1333698								
18	Eclosion hormone-1 (EH)	4284402-4284331	NA	x	NA	x	x	gi 108872486 gb EAT36711.1 	<i>A. aegypti</i>	7e-20
		4284135-4284052								
		4284002-4283847								
19	Eclosion hormone-2 (EH)	12709024-12708794	00178	x	x	x	x	gi 108869084 gb EAT33309.1 	<i>A. aegypti</i>	3e-08
		12708690-12708559								
20	FLRFa-like	265581-266078	07769	x	x	x	NA	gi 38424273 gb AAR19420.1 	<i>P. americana</i>	6e-13

Prohormones ^a	Scaffold Genbank accession number	ORF location	Glean	Prediction Algorithms			PSI- BLAST	CLUSTAL WPROF	Confirmation	BLAST	E- value	
				FGENESH	Augustus						Accession #	Organism
21	Insulin-like fragment	12552435-12552249	07035	x	x		x	NA	gi66770997 gb AAAY54810.1	<i>D. melanogaster</i>		1e-04
		12548565-12548490										
		12548443-12548344										
22	Insulin -like peptide	12521281-12521464	08479	x	x		x	NA	gi126335 sp PI5131 LIRP_LOCMI	<i>L. migratoria</i>		6e-10
23	IDLSPFYGHFNT-containing	12521512-12521720										
		2630460-2630594	04766	x	x		NA	NA	gi48102304 ref XP_392764.1	<i>A. mellifera</i>		1e-91
		2630677-2630962										
24	ITGQGNRIF-containing	2631016-2631209										
		15458008-15455709	08660	x	x		NA	NA	gi110762123 ref XP_001122204.1	<i>A. mellifera</i>		2e-53
		15455850-15455709										
25	SVDPIDGDLIG-containing	15455589-15455457										
		15454483-15454242	05944	x	NA		NA	NA		NA		NA
		15454158-15454024										
26	Myosuppressin	3775265-3775196										
		3774958-3774936	01469	x	NA		x	x	gi146019615 emb CAF04070.1	<i>B. germanica</i>		0.15
		3774866-3774356										
27	Neuroparsin	3774264-3774168	13258	x	x		x	x	gi14140213 emb CAC38869.1	<i>S. gregaria</i>		1e-07
		3774118-3774023										
		3773874-3773796										
28	Neuropeptide-Protein-like 1 (NPP1)	210887-211076										
		211207-211301	01469	x	NA		x	x				
		14447827-14447757	13258	x	x		x	x	gi14140213 emb CAC38869.1	<i>S. gregaria</i>		1e-07
29	Neurophlysin	14447709-14447582										
		14447532-14447402										
		15451371-15451828	06787	x	x		NA	NA	gi108868812 gb EAT33037.1	<i>A. aegypti</i>		6e-16
29	Neurophlysin	15451883-15452073										
		15452226-15452385										
		15452486-15452639										
29	Neurophlysin	12672392-12672270	NA	x	NA		x	NA	gi128109 sp P08163 NEUV_BUFA	<i>B. japonicus</i>		8e-27
		12672216-12672009										
		12671959-12671859										

Prohormones ^a	Scaffold Genbank accession number	ORF location	Glean	Prediction Algorithms			PSI-BLAST	CLUSTAL WPROF	Confirmation		E- value
				FGENESH	Augustus				Accession #	Organism	
30	Pheromone biosynthesis activating neuropeptide (PBAN)	11425007-11425070 11425121-11425207 11425253-11425554	08429	x	x		x	NA	gi1108878864 gb EAT43089.1	<i>A. aegypti</i>	0.018
31	Proctolin	4189506-4189578 4189335-4189231	15137	x	x		x	NA	gi24582626 ref NP_609158.2	<i>D. melanogaster</i>	0.010
32 a	Prothoracicotropic hormone (PTTH)/trunk	6122782-6123420	04893		x		x	NA	gi1118782122 ref XP_563293.2	<i>A. gambiae</i>	2e-44
32 b	Prothoracicotropic hormone (PTTH)	1106479-1106368 1106290-1106173 1104295-1104091	00735	x			x	x	gi1118788568 ref XP_555854.2	<i>A. gambiae</i>	5e-10
33	Pyrokinin/PBAN	1447393-1447312 1445073-1444888 1444829-1444609	04893	x	NA		x	x	gi466524 dbj BAA05954.1	<i>B. mori</i>	7e-07
34	Short neuropeptide F (sNPF)	5263880-5263778 5261779-5261624 5260268-5260222	NA	x	NA		x	NA		NA	NA
35	SIFamide	15595492-15595217	08645	x	NA		x	NA	gi1108873993 gb EAT38218.1	<i>A. aegypti</i>	1e-11
36	Sulfakinin (SK)	17019088-17019292 17019360-17019430	NA	x	NA		x	x	NA	NA	NA
37	Tachykinin (TK)	9074776-9074413 9074233-9073803	05685	x	x		x	x	gi86576812 gb AAX11211.2	<i>L. maderae</i>	1e-34

^aProhormones refer to precursor sequences for neuropeptides and hormones.