In silico identification of new secretory peptide genes in Drosophila melanogaster

LIU, Feng; Baggerman, G.; D'Hertog, W.; Verleyen, P.; Schoofs, L. & WETS, Geert (2006) In silico identification of new secretory peptide genes in Drosophila melanogaster. In: MOLECULAR & CELLULAR PROTEOMICS, 5(3). p. 510-522.

DOI: 10.1074/mcp.M400114-MCP200
Handle: http://hdl.handle.net/1942/1515

In silico identification of new secretory peptide genes in *Drosophila melanogaster*

Feng Liu*[1]

Geert Baggerman[1]

Wannes D'hertog

Liliane Schoofs

Geert Wets*

Laboratory for Developmental Physiology, Genomics and Proteomics, K.U.Leuven, Naamsestraat 59, B-3000 Leuven, Belgium.

* Data Analysis & Modeling Group, Faculty of Applied Economics, Limburg University,

University Campus - Building D, 3590 Diepenbeek, Belgium.

[1] These authors contributed equally to this work.

Corresponding author: geert.baggerman@bio.kuleuven.ac.be

Abstract

Bioactive peptides play critical roles in regulating most biological processes in animals. The elucidation of the amino acid sequence of these regulatory peptides is crucial for our understanding of animal physiology. Most of the (neuro)peptides currently known were identified by purification and subsequent amino acid sequencing. With the entire genome sequence of some animals now being available, it has become possible to predict novel putative peptides. In this way, BLAST analysis of the *Drosophila melanogaster* genome has allowed to annotate 36 secretory peptide genes so far. Peptide precursor genes are, however, very difficult to predict using BLAST, prompting us to search for an alternative approach described here. With the described program we have scanned the *Drosophila* genome for predicted proteins with the structural hallmarks of neuropeptide precursors. As a result, 119 putative secretory peptide genes were predicted, including the 43 annotated (neuro)peptides. These putative (neuro)peptide genes contain conserved motifs reminiscent of known neuropeptides from other animal species. Peptides that display sequence similarities to the mammalian vasopressin, atrial natriuretic peptide and prolactin precursors and the invertebrate peptides orcokin, PTTH, TMOF, DIMs among others were discovered. A peptidomic analysis of fruit fly hemolymph revealed the presence of two of the predicted peptides in this tissue confirming the viability of the genome screening method used.

Our data provides further evidence that many neuropeptide genes were already present in the ancestor of Protostomia and Deuterostomia, prior to their divergence. This bioinformatic study opens perspectives for the genome-wide analysis of peptide genes in other eukaryotic model organisms.

Key words: BLAST – peptide precursor gene – *Drosophila melanogaster* – *in silico* peptide identification – peptidomics - neuropeptide

Introduction

Peptides occur in the whole animal kingdom, from the least evolved phyla with a very simple nervous system (coelenterates) to the highest vertebrates. They play a key role in many if not all physiological processes as neurotransmitter, neuromodulator or neurohormone, and are therefore of considerable biological interest. Many dozens of peptides have already been discovered usually based on their biological activity. Peptides are synthesized in the cell in the form of large preproproteins, which are then cleaved and modified to generate biologically active peptides. Because of their critical signalling role, naturally occurring peptides play important roles in pathogenesis. Although they are of considerable biological, medical and industrial importance, a predictive method for the systematic identification of all candidate bioactive peptides in an organism is lacking so far. Expanding our knowledge on the structural level of peptides is, however, crucial in studying their role and interactions.

Computational methods have become especially important since the advent of genome projects. By means of BLAST (Basic Local Alignment Searching Tool) analysis, an organism genome can be screened for peptide encoding genes, based on sequence similarity to known peptide genes from other organisms. Using BLAST, 36 peptide genes have already been found in *Drosophila melanogaster* (1,2,3,4). Likewise in *Anopheles gambiae*, 35 peptide-encoding genes were discovered using the same sequence similarity-based mining approach (5). While certain of these peptides have been studied in detail (6,7,8), more data on the entire peptidome of insects are needed for integrated functional analyses.

However, for *in silico* prediction of peptide precursor genes in large sequence datasets, the performance of the BLAST tool is limited because putative peptide sequences for which no orthologous (similar) biologically active peptide has been identified as yet (for instance, because of lack of suitable detection methods) will not be revealed. BLAST programs are based on similarity between sequences and do not take into account the existence of other structural hallmarks, in this case those of

peptide precursors. In addition, while BLAST programs are very suitable to scan databases for conserved proteins (9), they are far less efficient at finding similarity to short peptides when they are scanned against the whole genome sequence. Indeed, in most cases, only a short conserved motif is responsible for the function of a particular peptide and often only this short sequence motif, which can be 5 amino acids or less in length, is conserved. For instance, members of the invertebrate FMRFamide peptide family can share the carboxyterminal tetramer FMRFamide or the MRFamide tripeptide motif or only the RFamide motif.

In recent studies of Baggerman et al. (10,11,12), the peptidome (the battery of all present peptides) of the larval *Drosophila* central nervous system was analysed at the amino acid sequence level by means of nanoscale liquid chromatography combined by tandem mass spectrometry and database mining. These results provided biochemical evidence for the expression of not less than 40 peptides in the *Drosophila* brain at a specific time point. Interestingly, not only known or predicted peptides were identified in this study but also eight additional peptides that are encoded in 5 novel peptide genes that could not be identified by BLAST and hence were not predicted as being peptide precursor genes. Unfortunately, several as yet unknown peptide ions observed in the same study could not be fully sequenced. The lack of an appropriate database of putative peptides could be the reason. Sequencing of peptides is aided a great deal if an appropriate database resource is available. Naturally occurring (neuro)peptides are synthesized in the cell as protein precursors from which the peptides are cleaved by specific enzymes. The problem with the identification of these peptides by mass spectrometry combined with database searching (peptidomics) is that many of the cleaving enzymes are not known or that their specificity is not fully understood. This means that in a Mascot search the cleaving enzyme or method cannot be specified. In addition, peptides can have a fairly large number of PTMs (amidation, pyroglutamic acid, …). All this puts extra strains on the database searching. Therefore, reducing the size of the protein

database by including only putative peptide precursors makes the identification by peptidomics easier and reduces the computing time.

The bioinformatic approach described here will complement direct peptidome experiments. Peptidome experiments do not take into account the structural attributes of peptide precursors, such as the presence of a signal peptide, mono-and dibasic cleavage sites, etc.. Therefore, it is highly likely that with a peptidome experiment many false positives are picked up (especially when using inappropriate extraction methods), as reported recently (13). Another disadvantage of peptidome experiments is that (i) not all peptides are equally well extracted (ii) not all peptides ionise with the same efficiency and (iii) not all peptides are present in the same concentration. All prompted us to construct a peptide database in order to provide a strong support for future peptide research. In this study we describe a new in silico searching program that employs typical hallmarks of biological peptides and their precursors, which are not used by the currently available predictive algorithms.

Experimental procedures

Rationale

First, we examined the structural hallmarks of known peptide precursor proteins. Regulatory peptides are synthesized as part of larger precursor proteins that are subsequently processed into smaller active substances. All peptide precursor proteins are usually less than 500 amino acids in length and contain an N-terminal signal peptide (corresponding to a transmembrane domain) that directs them into the secretory pathway of the cell. After cleavage of the signal peptide, further processing by endoproteases occurs predominantly at processing sites, typically mono- and dibasic amino acid residues (14). For example, insect antimicrobial peptides (like defensins) contain multiple RK, RR, KK repeats within their sequences (15). Many peptide precursors contain multiple bioactive peptides that are often highly related,

for example the tachykinin precursor, the allatostatin precursors and the neuropeptide F precursor in *Drosophila* and other insects (16,17,18,19,20). However, in *Drosophila* as in other insects, peptide genes may also encode multiple, unrelated bioactive peptides or just a single bioactive peptide (21). Based on the common structural characteristics of known invertebrate peptide precursors we build a sensitive searching procedure to identify peptide genes in the *Drosophila* database. Two types of programs were constructed. The first one was built to find those peptide precursors that encode multiple highly related peptides. The second one searches for precursors containing a single peptide or multiple unrelated peptides that share conserved motifs with known peptide precursor proteins from other animal species. In both cases, the putative peptide sequence was defined by the presence of characteristic proteolytic cleavage sites, flanking the peptide sequences.

In order to avoid the shortcomings of BLAST programs in searching long sequences for short similarity, we split the protein sequences from *D. melanogaster* into short subsequences (in silico cleavage at their putative processing sites), and then applied BLAST to compare the subsequences within each protein sequence to fish for those precursor sequences that have at least 2 similar subsequences. Additionally, the subsequences are also compared with subsequences derived from known peptide precursors in the Swissprot database. These subsequences are obtained by in silico processing the known peptide precursors from other animal species obtained from the Swissprot database. Because each *Drosophila* protein sequence is split into a number of subsequences, and because all of these subsequences are subsequently compared with all known peptide precursor subsequences, a very large number of alignments with a high score are obtained. Because similarity does not imply homology, only the alignments with sequence motifs from actual bioactive peptides are considered significant and the obtained subsequences as possible peptides. A fasta protein database containing all identified putative peptide precursors was constructed. This database was loaded on an in house Mascot server and used for

the identification of peptides in a peptidomic analysis of *Drosophila* hemolymph. Our results show that this technique is very efficient to find novel peptide genes.

The program.

The aim of the program is to mine for putative peptide precursors according to the rules and the techniques described above. The program is implemented in SAS, a powerful integrated software to access, manage, analyse and present data. External tools such as SignalP and BLAST need to be run independently. They communicate with the program by text files. The program includes a few sub-programs listed below.

Protein.SAS. The first part of the program, named Protein.SAS, serves to pick up all proteins from a specific species, in this case *Drosophila melanogaster*. The input of the sub-program consists of the Swissprot protein database files, and additional *Drosophila* genes at GenBank identified by Hild et al. (2003) (21). The relevant information for each of the *Drosophila* proteins, such as accession number, protein name, gene name, protein sequence, signal peptide information, length and mass is written into a SAS dataset. The first 70 amino acids of every protein sequence serve as output to a text-file in FASTA format, which is used as the input of SignalP. SignalP (www.cbs.dtu.dk/services/SignalP) for eukaryotes is then run to predict the presence and location of a signal peptide in each protein sequence. Next, the sub-program reads the output file by SignalP, and another SAS dataset is created which includes the predicted signal information of every *Drosophila* protein. The dataset is compared with the dataset of all *Drosophila* proteins, and the proteins are retained if they are either annotated to have signal peptides in the Swissprot protein database files or predicted to have signal peptides by SignalP. The comparison result is a dataset of *Drosophila* proteins having N-terminal signal peptides. From this dataset, only the proteins which are less than 500 amino acids in length are retained. In total,

5096 proteins make up the final *Drosophila* protein dataset which will be analysed further. The logic of Protein.SAS is illustrated in Figure 3.

Peptide.SAS. The sub-program serves to filter all the peptides or their precursors that are known in Metazoa today from Swissprot protein databases. The peptide precursors are identified by the keywords in each protein datafile. If a protein is picked up, its relevant information, like the information collected in *Drosophila* proteins, is written into a SAS dataset. Figure 4 describes the process.

Cleavage.SAS. The objective of the sub-program is to split protein sequences into subsequences after removal of the signal peptide sequence. The protein sequences are split *in silico* at cleavage sites, typical for peptide precursors. Conventional amino acid motifs that are required for cleavage of neuropeptides from their protein precursors in insects have been described as: GKR, GRR, GR, GRK, GKK, KR, RR, GK, RK, KK, R (22). From our statistical analyses on all known peptide precursors in all organisms (data not shown), it is clear that the processing of peptide precursors does not occur at every conventional cleavage site in the precursor. Cleavage also depends on the amino acids that are at the proximity of the cleavage site. For example, proteolytic processing at GKK followed by R always occurs. However, if GKK is followed by A, N, S or K, the processing may or may not occur. For other amino acids at this position, it has not been demonstrated whether processing occurs. Second example: proteolytic processing at a single R residue only occurs when there is a basic amino acid residue in position -4, -6 or –8 with respect to the single R. The basic amino acid is usually an R, but K or H residues work as well [19].

BLAST analysis. The output of Cleavage.SAS consists of two database files: a database of "*Drosophila* subsequences" and a 'peptide' database of "known metazoan peptides". BLAST analysis is then conducted on these two databases.

The score matrix **PAM30** is used, and the expectation value (e-value) as well as the parameter **word size** are set to 6 and 2 respectively in order to find short but strong similarities. Figure 5 explains the process.

Extract.SAS, Shift.SAS, Motif.SAS. These programs are used to screen the result output by BLAST and determine the biologically significant matches. The sub-program Extract.SAS extracts the *Drosophila* proteins which have at least two similar subsequences within the protein. The sub-program Shift.SAS reads the comparison result from the BLAST analysis and computes the shift value. The shift value is the minimal distance between the N- or C- terminal of a subsequence and the matching amino acids in the subsequence. From the statistical analysis of the known peptide precursors, these shift values should be low. This means that the motif should be close to a cleavage site. The shift value is set to be no larger than 3 in the program. The sub-program Motif.SAS reads the comparison results between *Drosophila* subsequences and known metazoan peptide subsequences, as well as the comparison results among known metazoan peptide sequences themselves, and identifies the *Drosophila* subsequences that contain conserved peptide motifs.

Tmpred and SOSUI. Finally, online software TMpred at http://www.ch.embnet.org/software/TMPRED_form.html (23) and SOSUI http://sosui.proteome.bio.tuat.ac.jp/sosuimenu().html are used to determine whether a protein has a single transmembrane region at its N-terminus (Fig. 6). The minimum and maximum length of the hydrophobic part of the transmembrane region was respectively set at 17 residues and 33 residues respectively. For the TMpred program a score above 500 for both inside to outside as well as outside to inside helices was considered to be significant for the presence of the N-terminal transmembrane region. A score of 250 was considered to be significant for the presence of an inside to outside helix of any second or third transmembrane region.

A putative peptide precursor was retained if one program predicts a single transmembrane region at the N-terminus. When both programs predict the absence of an N-terminal transmembrane region, the protein was deleted from the list. The cut-of of the start of the transmembrane region was set at the 20th residue; transmembrane regions that started at or after this point were not considered to be at the N-terminal side and corresponding proteins were deleted from the database.

Mass spectrometry

Animals: *D. melanogaster* are kept in 250 ml bottles and bread on a standard diet that consists of 70 ml water; 17 g sucrose; 0.45 g yeast; 0.9 g agar; 0.5 ml 8 % Nipagin and 0.36 ml propionic acid. Wandering stage larvae were collected. The larvae were washed in water and their cuticle was punctured with a fine stainless steel needle. Hemolymph leaking from the wound was collected using glass microcaplilaries and transferred to a 0.5 µl micro tube containing 200 µl of ice cold methanol/water/acetic acid (90/9/1, v/v/v). In total 30 µl of hemolymph was collected. After the extraction the sample was centrifuged, the supernatant was dried, dissolved in 15 µl of 2 % acetonitrile in MQ-water with 0.1% formic acid and filtered. Ten µl of this sample was injedted on the LC-Qtof MS/MS system

Capillary LC-tandem MS experiments were conducted using an Ultimate HPLC pump, a column-switching device (Switchos) and a Famos autosampler (all LC Packings, The Netherlands) coupled to a Q-Tof mass spectrometer (Micromass, UK). Chromatography was performed using a guard column (µ-guard column MGU-30 C18, LC-Packings, The Netherlands) acting as a reverse phase support to trap the peptides. Ten µl of the sample (corresponding to 50 *Drosophila* CNS equivalents) was loaded on the pre-column with an isocratic flow of 2 % acetonitrile in MQ-water with 0.1% formic acid at a flow rate of 10 µl/min. After 2 min, the column-switching valve was switched, placing the pre-column online with the analytical capillary column, a Pepmap C18, 3µm 75µm x 150mm nano column (LC Packings, The

Netherlands). Separation was conducted using a linear gradient from 95% solvent A, 5% solvent B to 80% A, 20% B in 90 minutes, followed by a linear gradient from 80% A, 20% B to 50% A, 50% B in 60 minutes (solvent A: water, formic acid; 99.9/0.1 (v/v); solvent B: acetonitrile, formic acid; 99.9, 0.1 (v,v)). The flow rate was set at 150 nl/min.

The LC system was connected in series to the electrospray interface of the Q-Tof device. The column eluent was directed through a stainless steel emitter (Proteon, Denmark). Needle voltage was set at 1,650 Volts, cone voltage at 35 Volts. Nitrogen was used as nebulising gas. Parent ions with 2, 3 or 4 charges of sufficient ion intensity (threshold was set at 15 counts/second) were automatically recognized by the charge state recognition software (MassLynx 3.5, Micromass, UK) and selected for fragmentation as they elute from the column. Argon was used as a collision gas; collision energy was set at 25 to 40 eV depending on the mass and charge state of the selected ion. The detection window in the survey scan was set from 400 to 1400 mass to charge (m/z). Fragmentation spectra were acquired from m/z 50 to 2000.

Identification of peptides.

Peptides were identified using Mascot (Matrixscience, UK). The database used for these searches was the database of putative secretory peptide precursors from *Drosophila.* The setting were as follows: peptide tolerance was set at ± 0.3 Da, MS/MS tolerance at ± 0.2. Enzyme was set to none and variable modifications amide (C-terminal), pyroglutamine (Q) and oxidation (M) were selected. Only hits with significance level of 95% or higher were retained.

Results

1. Construction of two databases.

First, we constructed two databases. The first database is generated as follows: *Drosophila* proteins that are less than 500 amino acids in length and that start with a

signal peptide are assembled from SWISS and TrEMBL databases, as well as from a collection of additional *Drosophila* genes identified by Hild et al. (22). The program SignalP for eukaryotes was used to predict the occurrence of a signal peptide for a protein sequence (24). As a result, 5096 *Drosophila* protein sequences are retained. Then, all these protein sequences were split into short subsequences at the conventional cleavage sites, taken into account the nature of the amino acids in the proximity of each cleavage site. These subsequences form the first database, which we named '*Drosophila* subsequence'.

The second database is a 'peptide' database that comprises the subsequences - obtained by in silico cleavage at mono-or dibasic processing sites- of all known peptide precursor proteins known in metazoans to date. These annotated peptides or peptide precursors were filtered from the SWISS_PROT (release 42.11) and TrEMBL (release 25.11) databases as follows: A protein is retained when it is annotated as (neuro)peptide precursor or when its name contains the word 'neuropeptide'. Proteins of which the corresponding protein file contains keywords such as peptide, neuropeptide, hormone or neurotransmitter are also retained. But if these proteins have a subcellular location as membrane protein (as indicated in the protein file) or if they are characterised by key words such as receptor, signal-anchor, transmembrane, binding protein, DNA binding, nuclear protein, nuclear transport, enzyme or words ending in 'ase', they are excluded. In total, 2858 proteins meet these criteria. These peptide precursors are subsequently split into short subsequences at the conventional cleavage sites, also taken into account the character of the amino acids in the proximity of each cleavage site. This collection of peptide precursor subsequences constitutes the 'peptide' database.

2. Setup of datamining analysis

Standalone BLAST is used to compare the two above-mentioned databases. Interpretation of the results generated by BLAST involves evaluation of the matches

to determine whether they are significant. Therefore, genuine and biologically meaningful similarities need to be distinguished from the irrelevant and essentially random ones. If the alignment is similar to a motif, it is considered significant, and the subsequence is considered a putative peptide. In order to find the conserved motifs, all known peptide precursor subsequences were compared by BLAST.

Four types of analysis were performed:

1. The '*Drosophila* subsequence' database is compared with itself and those protein sequences, which have at least two similar subsequences within the same protein sequence, are retained (first screening method)

2. The peptide precursor subsequences in the 'peptide' database are compared with each other and the obtained similar amino acid sequence tags are considered as possible motifs.

3. The '*Drosophila* subsequence' database is compared with the 'peptide' database and those *Drosophila* subsequences that display sequence similarities to a conserved motif within a known peptide precursor subsequence in another metazoan organism are retained (second screening method).

4. The retained proteins from 1 and 3 are then analysed by a transmembrane prediction method. Our analysis of the annotated *Drosophila* neuropeptide precursors indicates that almost all have a single transmembrane region, which is located at the N-terminus and which corresponds to the signal peptide Therefore, in a final step we fine-tuned the generated list of putative peptide precursors based on this hallmark. The list was curated by the deletion of (i) all soluble proteins (lacking membrane-spanning regions), (ii) of proteins having more than one transmembrane region and (iii) of proteins having one transmembrane region that is not located in the N-terminal region.

Screening method 1

The first screening method is based on the principle that multiple peptides encoded by a single invertebrate peptide precursor gene are often highly related. Therefore, proteins were only selected if they have at least 2 similar subsequences and if the matching amino acid sequence is at or close to the N or C terminus of at least one subsequence.

Therefore, the structural pattern of a putative peptide precursor is:

……[cleavage1]-x1(3,60)-[cleavage2]-……-[cleavage3]-x2(3,60)-[cleavage4]…….

x1(3,60) and x2(3,60) are two similar subsequences which are between 3 and 60 amino acids long. [cleavage1(-4)] can be any conventional cleavage site listed above. The subsequences do not need to be adjacent within the precursor.

Using this screening method we found 58 peptide precursors in *Drosophila*, 10 of which are well known peptide precursor genes that encode at least 2 related bioactive peptides, drosulfakinin (dsk), FMRFamide, shortNPF, tachykinin, capa or mt-cap2b, diuretic hormone.

For example, the protein identified by accession number Q9V808 is a putative peptide precursor. By comparing database '*Drosophila* subsequence' with itself, we obtained three similar subsequences in Q9V808 (see example in Fig.5). All the putative *Drosophila* peptide precursors, mined by this screening method are depicted in Table 1.

Screening method 2.

The fact that only 9 of the 44 known neuropeptide precursors as well as one known immune induced peptide in *Drosophila* were listed by the first screening method, indicated that the catalogue of putative regulatory peptide precursors obtained by the first screening method is doubtless incomplete. Therefore, we set out for a second screening method that screens for *Drosophila* proteins having a signal peptide and of which at least one subsequence has at least 3/5 amino acids at or close to the N or C

terminus identical to a known peptide. In addition, the identical 3/5 amino acids should be similar to a conserved motif present in known peptides. The retained proteins are then further filtered by the transmembrane prediction analysis as in the first method.

By means of the second method we found 70 *Drosophila* peptide precursor genes in total, 42 of which are known peptide precursors and 28 are novel. Each of these putative peptide precursor genes encodes multiple non-related peptides or only a single putative peptide. For example, protein Q8MS86 was identified as a putative peptide precursor. The similar subsequence is Q8MS86_2: WKILTAGSHFRWL. The similar known peptides are P11885_2: YVMSHFRWNKF from *Rana catesbeiana*, and P06298_8: NGNYRMHHFRWGSPPKD from *Xenopus laevis*. The total output of putative peptide precursors mined by this screening method is shown in Table 2.

The combined computational methods generated in total 75 novel putative peptide precursors in *D. melanogaster*, in addition to the 43 known ones.

Peptidomic analysis

The nanoLC-tandem MS method allows us to select and fragment the peptide ions as they elute from the column, even when co-eluting with other peptides. Peptides were identified by subjecting their fragmentation spectra to a Mascot search on an in-house server. This bioinformatics tool (http://www.matrixscience.com) allows the identification of proteins and peptides by matching MS data against any FASTA format protein or (translated) nucleic acid sequence database. In a typical MS/MS ion search, we combined all MS/MS data of every peptide selected for fragmentation during a LC-MS run, in a comprehensive peak list. This type of file contains the centroided mass values and associated intensity values of all the parent ions selected and corresponding fragmentation peaks, and can be submitted to Mascot for fully automated identification of several tens of peptides at the same time.

In total more than 500 ions were automatically selected for fragmentation. Twenty peptides were identified most of which are known to occur in the hemolymph such as the Attacins and DIMs (14). In addition we identified 2 novel peptides: LDDSENNDQVVGLLDVADQGANHANDGAREA and a truncated form of this peptide, LLDVADQGANHANDGAREA (Fig. 7). These peptides originate from protein CG7738 which was identified as a putative peptide precursor by screening method 1 (Fig. 8).

Discussion

Because of the availability of its complete genome sequence, *Drosophila* becomes a model insect for peptide research. We have identified in total 118 putative peptide precursor genes in *D. melanogaster* by applying the here presented database searching programs. 43 of them are annotated peptide precursors. All predicted peptide precursors meet with following criteria: (i) each putative peptide precursor is less than 500 amino acids in length and has a signal peptide; (ii) each precursor contains one or several putative peptides that are flanked by conventional cleavage sites. Here are two possibilities: the precursor contains two or more peptides that share sequence similarities or alternatively, the precursor contains a single peptide that shares conserved motifs with known peptide precursor subsequences from other organisms; (iii) all predicted peptide precursors have one N-terminal transmembrane region.

Several of the genes mined by our method encode peptides that display significant sequence similarities to known vertebrate or invertebrate neuropeptides. These similarities have not been discovered by BLAST scanning of the whole *Drosophila* genome. We will discus a few examples. A putative peptide encoded by CG3868, mined by the first method, displays sequence similarities with a antifreeze glycopeptide precursor identified in Antarctic fish (26). The salivary gland glue protein (CG18087) contains a putative peptide sequence that displays significant

similarities to vertebrate neurophysins. Neurophysins are a group of small, soluble proteins secreted by the hypothalamus. They serve as binding proteins for oxytocin and vasopressin during their transport to the posterior pituitary. They are secreted with the hormones but have no known functions other than serving as a carrier. In vertebrates neurophysins originate from the vasopressin peptide precursor. The salivary gland glue protein (CG18087) does not contain a vasopressin/oxytocine-like peptide.

Putative peptides from two genes, CG9358 and BK003312, display sequence similarities to conserved parts of the prolactin precursor. Prolactin and growth hormone are two distinct neuropeptide hormones that have been found in all vertebrate groups but not in cyclostomes (27), although prolactinergic neurons that were detected immunochemically occur in a protochordate (28). The GH/PRL superfamily is likely to have a prevertebrate origin but a putative invertebrate member was so far not found, in contrast to other neuropeptide superfamilies that are highly conserved in vertebrates and invertebrates. Examples are tachykinins, gastrin, insulin, neuropeptide Y, corticotropin releasing factor, calcitonin-gene related peptide (29).

Drosophila BK002187 encodes a peptide with sequence similarities to atrial natriuretic peptide (ANP). Natriuretic peptides are vertebrate hormones that play a pivotal role in cardiovascular and body fluid homeostasis in vertebrates (30). Although a novel natriuretic peptide has recently been found in the heart and brain of the hagfish, the most primitive vertebrate (31), no member of this family was has as yet been described in invertebrates. Finally, a peptide encoded by the *Drosophila* LP04693 displays sequence similarities to γ-MSH, a pituitary hormone, derived from the pro-opiomelanocortin precursor, the function of which has remained elusive (32) (Fig. 7)

When we consider similarities to invertebrate neuropeptides that are known (annotated) at this moment, the CG1565 protein contains a putative peptide that has

an N-terminal hexamer, contained within orcokinin, a myotropic neuropeptide discovered in crustaceans (33,34), but which has so far not been identified in insects (Fig. 8). A putative peptide sequence within the trunk protein precursor displays striking similarities with prothoracicotropic hormone, a neuropeptide that has so far only been identified in lepidopteran species, in which it stimulates ecdyson biosynthesis in the prothoracic glands (35). Interestingly, the sex-specific gene, MSOPA, as identified by Jin et al. (36) encodes a putative peptide that shows sequence similarities to a male accessory gland-specific 57kDa peptide precursor. Next, the putative *Drosophila* peptide encoded by CG8087 displays more than 60% sequence identities with a neuropeptide derived from a neurospecific peptide precursor in the terrestrial snail, *Helix lucorum* (37). Finally, some mined genes (CG16882, CG11131, CG7465, CG1221, Argos, Trunk) have been predicted or shown to encode for ligands of membrane receptors, such as EGF, Toll or Torso receptors (38), a function in line with the peptidergic nature of their products.

Since the publication of the *Drosophila* genome sequence, several microarray studies have been performed and we observed that some of the mined peptide precursor genes are upregulated by ecdyson (CG7350, CG7608, CG1807, CG 7350) (39,40). Ecdyson is an ecdysteroid involved insect metamorphosis and reproduction. It is the precursor of 20-OH-ecdysone, the functional counterpart of vertebrate estrogen (41). In this way, our data are in accordance with the reported interactions of peptide and steroid hormone signalling cascades in vertebrates (42,43). Other mined genes are upregulated after infection (44) and encode for peptides that are secreted into the haemolymph such as attacin, diptericin, drosocin and various *Drosophila* immune induced peptides or DIMs (15). With the currently established program, several additional putative peptide precursor genes that display sequence similarities to known DIMs, were found. Three of them (CG32851, CG5791, CG15065) form part of the Toll pathway (45) and one (CG18107) is rhythmically expressed in the head (46)

Former microarray studies revealed that regulation of transcription of known neuropeptide genes as well as other putative peptide precursor genes established in this study is circadian clock dependent (i.e. capa, corazonin, CG4784, CG1807; (46,47), nutrient-dependent (CG10918, CG15225) (48) or sex-specific (accessory gland peptides, CG7738, CG11458) (36,49).

Our program also picked up the drosocrystallin gene (50) as well as other annotated cuticular proteins. In *Tenebrio molitor*, biologically active peptides display strong sequence similarities to parts of cuticle proteins and therefore they might be processed from them (51).

Given the fact that proteolytic processing does not always occur at every conventional cleavage site (52), our established catalogue of predicted peptide precursors is doubtless incomplete and it will be a difficult challenge to consider the existence of these unconventional cleavage sites in the further refinement of our method.

Only two of the characterized peptide precursors were not mined by our method, i.e. the diuretic hormone precursor or CG8348 because it has 4 transmembrane regions, and the proctolin precursor 'Q8MMJ7' because it sequence is too short (5 amino acids) to be filtered by the program. Inherent to datamining methods, a few cases could be false positives: CG5559 has been annotated to encode a conserved protein involved in synaptic vesicle fusion, CG6409 has been predicted to be a component of the endoplasmatic reticulum, CG11577 has been predicted to permanently reside in the lumen based on its C-terminal sequence (53) and CG6357 encodes a putative cysteine protease.

The database of predicted and known peptide precursors in *Drosophila* as established in this paper will serve several applications in experimental research. Many unassigned masses observed in peptidomic experiments could not be identified, which could be attributed to the lack of an appropriate peptide database (10). Mass spectrometric data will become much easier to read and interpret if the

database against which they are scanned is much smaller than the SWISS-PROT database. In a peptidomic analysis of the hemolymph of the fruit fly, we were able to identify 2 novel peptides originating from the CG7738. The first peptide, LDDSENNDQVVGLLDVADQGANHANDGAREA is 31 AA in length and is flanked at the amino-terminal side by the cleavage site of the signal peptide (Fig. 8). At the carboxyterminus, the peptide is flanked by an arginine residue that could act as a monobasic cleavage site. The second peptide, LLDVADQGANHANDGAREA, is a truncated homologue of the first one. This example clearly demonstrates that the peptide database identified in this study facilitates the mass spectrometric identification of peptides in *Drosophila*.

Also in mammalian models, genome-wide analysis of peptides by mass spectrometry has recently boosted (54). Construction of a peptide database, like the one presented here for *Drosophila*, will be of high value to support these studies. As the structural hallmarks of peptide precursor sequences are highly conserved across phyla, we foresee that the established search program can be adapted for the genome-wide analysis for peptide precursor genes in other animal model systems that have a sequenced genome.

Acknowledgements.

This research was sponsored by the FWO grant G0146.03. G. Baggerman is a post-doc fellow of the FWO.

References

1. Vanden Broeck, J. (2001) Neuropeptides and their precursors in the fruitfly, *Drosophila melanogaster*. *Peptides* 22,241-254.

2. Hewes, R. S. and Taghert, P. H. (2001) Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. *Genome Res.* 11,1126-1142.

3. De Loof, A., Baggerman, G., Breuer, M., Claeys, I., Cerstiaens, A., Clynen, E., Janssen, T., Schoofs, L., and Vanden Broeck, J. (2001) Gonadotropins in insects: an overview. *Arch. Insect Biochem. Physiol* 47,129-138.

4. Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001) An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* 11,213-221.

5. Riehle, M. A., Garczynski, S. F., Crim, J. W., Hill, C. A., and Brown, M. R. (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* 298,172-175.

6. Gade, G., Hoffmann, K. H., and Spring, J. H. (1997) Hormonal regulation in insects: facts, gaps, and future directions. *Physiol. Rev.* 77,963-1032.

7. Gade, G. (2004) Regulation of intermediary metabolism and water balance of insects by neuropeptides. *Annu. Rev. Entomol.* 49,93-113.

8. Nassel, D. R. (2002) Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog. Neurobiol.* 68,1-84.

9. Yona, G. and Levitt, M. (2002) Within the twilight zone: a sensitive profile-profile comparison tool based on information theory. *J. Molec. Biol.* 315,1257-1275.

10. Baggerman, G., Cerstiaens, A., De Loof, A., and Schoofs, L. (2002) Peptidomics of the larval *Drosophila melanogaster* central nervous system. *J. Biol. Chem*. 277,40368-40374.

11. Baggerman, G., Boonen, K., Verleyen, P., De Loof, A., and Schoofs, L. (2004) Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two- dimensional capillary liquid chromatography Q-Tof mass spectrometry. *J. Mass Spectrom*. 40,250-260,

12. Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., De Loof, A., and Schoofs, L. (2004) Peptidomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 803,3-16.

13. Skold, K., Svensson, M., Kaplan, A., Bjorkesten, L., Astrom, J., Andren, P.E (2002) A neuroproteomic approach to targeting neuropeptides in the brain. *Proteomics* 2,447-54.

14. Canaff, L., Bennett, H. P., and Hendy, G. N. (1999) Peptide hormone precursor processing: getting sorted? *Mol. Cell Endocrinol*. 156,1-6.

15. Uttenweiler-Joseph, S., Moniatte, M., Lagueux, M., Van Dorsselaer, A., Hoffmann, J. A., and Bulet, P. (1998) Differential display of peptides induced during the immune response of *Drosophila*: a matrix-assisted laser desorption ionization time-of- flight mass spectrometry study. *Proc. Natl. Acad. Sci. U.S.A* 95,11342-11347.

16. Nichols, R., Schneuwly, S. A., and Dixon, J. E. (1988) Identification and characterization of a *Drosophila* homologue to the vertebrate neuropeptide cholecystokinin. J.*Biol.Chem.* 263,12167-12170.

17. Nichols, R., Bendena, W. G., and Tobe, S. S. (2002) Myotropic peptides in *Drosophila melanogaster* and the genes that encode them. *J. Neurogenet.* 16,1-28.

18. Schneider, L. E. and Taghert, P. H. (1988) Isolation and characterization of a *Drosophila* gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH2 (FMRFamide). *Proc. Natl. Acad. Sci. U.S.A* 85,1993-1997.

19. Siviter, R. J., Coast, G. M., Winther, A. M., Nachman, R. J., Taylor, C. A., Shirras, A. D., Coates, D., Isaac, R. E., and Nassel, D. R. (2000) Expression and functional characterization of a *Drosophila* neuropeptide precursor with homology to mammalian preprotachykinin A. *J. Biol. Chem*. 275,23273-23280.

20. Vanden Broeck, J., Veelaert, D., Bendena, W. G., Tobe, S. S., and De Loof, A. (1996) Molecular cloning of the precursor cDNA for schistostatins, locust allatostatin-like peptides with myoinhibiting properties. *Mol. Cell Endocrinol.* 122,191-198.

21. Taylor, C. A., Winther, A. M., Siviter, R. J., Shirras, A. D., Isaac, R. E., and Nassel, D. R. (2004) Identification of a proctolin preprohormone gene (Proct) of *Drosophila melanogaster*: expression and predicted prohormone processing. *J. Neurobiol*. 58,379-391.

22. Hild, M., Beckman, B., Haas, S., Koch, B., Solvyev, V., Busold, C., Feelenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J., and Paro, R. (2003) An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biology* 5,R3.

23. Veenstra, J. A. (2000) Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch. Insect Biochem. Physiol* 43,49-63.

24. Hofmann, K. and Stoffel, W. (2004) Tmbase - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* 374,166.

25. Nielsen, H., Brunak, S., and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* 12,3-9.

26. Osuga, D. T. and Feeney, R. E. (1978) Antifreeze glycoproteins from Arctic fish. *J. Biol. Chem*. 253,5338-5343.

27. Forsyth, I. A. and Wallis, M. (2002) Growth hormone and prolactin – molecular and functional evolution. J*. Mammary Gland. Biol. Neoplasia* 7,291-312.

28. Pestarino, M. (1983) Prolactinergic neurons in a protochordate. *Cell Tissue Res.* 233,471-474.

29. De Loof, A. and Schoofs, L. (1990) Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources. *Comp. Biochem..Physiol* B 95,459-468.

30. Takei, Y. (2000) Structural and functional evolution of the natriuretic peptide system in vertebrates*. Int..Rev..Cytol*. 194,1-66.

31. Kawakoshi, A., Hyodo, S., Yasuda, A., and Takei, Y. (2003) A single and novel natriuretic peptide is expressed in the heart and brain of the most primitive vertebrate, the hagfish (*Eptatretus burgeri*). *J. Mol. Endocrinol.* 31,209-220.

32. Humphreys, M. H. (2004) Gamma-MSH, sodium metabolism, and salt-sensitive hypertension. *Am. J. Physiol Regul. Integr. Comp Physiol* 286,R417-R430.

33. Stangier, J., Hilbich, C., Burdzik, S., and Keller, R. (1992) Orcokinin: a novel myotropic peptide from the nervous system of the crayfish, *Orconectes limosus*. *Peptides* 13,859-864.

34. Huybrechts, J., Nusbaum, M. P., Bosch, L. V., Baggerman, G., De Loof, A., and Schoofs, L. (2003) Neuropeptidomic analysis of the brain and thoracic ganglion from the Jonah crab, *Cancer borealis*. *Biochem. Biophys. Res .Commun.* 308,535-544.

35. Xu, W. H., Rinehart, J. P., and Denlinger, D. L. (2003) Structural characterization and expression analysis of prothoracicotropic

hormone in the corn earworm, *Helicoverpa zea*. *Peptides* 24,1319-1325.

36. Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* 29,389-395.

37. Bogdanov, Y., Balaban, P. M., Zakharov, I. S., Poteryaev, D. A., and Belyavsky, A. V. (1996) Identification of two novel genes specifically expressed in the D-group neurons of the terrestrial snail CNS. *Invert. Neurosci.* 2,61-69.

38. Casali, A. and Casanova, J. (2001) The spatial control of Torso RTK activation: a C-terminal fragment of the Trunk protein acts as a signal for Torso receptor in the *Drosophila* embryo. *Development* 128,1709-1715.

39. Fletcher, J. C. and Thummel, C. S. (1995) The Drosophila E74 gene is required for the proper stage- and tissue-specific transcription of ecdysone-regulated genes at the onset of metamorphosis. *Development* 121,1411-1421.

40. Lehman, M. (2003) *Drosophila* Sgs genes: stage and tissue specificity of hormone responsiveness. *Bioessays* 18,47-54.

41. De Loof, A. and Huybrechts, R. (1998) "Insects do not have sex hormones": a myth? *Gen. Comp Endocrinol.* 111,245-260.

42. Richards, J. S., Russell, D. L., Ochsner, S., Hsieh, M., Doyle, K. H., Falender, A. E., Lo, Y. K., and Sharma, S. C. (2002) Novel signaling pathways that control ovarian follicular development, ovulation, and luteinization. *Recent Prog. Horm. Res.* 57,195-220.

43. Cardona-Gomez, G. P., Mendez, P., DonCarlos, L. L., Azcoitia, I., and Garcia-Segura, L. M. (2002) Interactions of estrogen and insulin-like growth factor-I in the brain: molecular mechanisms and functional implications. *J.Steroid Biochem. Mol. Biol.* 83,211-217.

44. De Gregorio, E., Spellman, P. T., Rubin, G. M., and Lemaitre, B. (2001) Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proc.Natl.Acad.Sci.U.S.A* 98,12590-12595.

45. Boutros, M., Agaisse, H., and Perrimon, N. (2002) Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev.Cell* 3,711-722.

46. McDonald, M. J. and Rosbash, M. (2001) Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* 107,567-578.

47. Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N., and Young, M. W. (2001) Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron* 32,657-671.

48. Zinke, I., Schutz, C. S., Katzenberger, J. D., Bauer, M., and Pankratz, M. J. (2002) Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J.* 21,6162-6173.

49. Fujii, S. and Amrein, H. (2002) Genes expressed in the *Drosophila* head reveal a role for fat cells in sex-specific physiology. *EMBO J.* 21,5353-5363.

50. Janssens, H. and Gehring, W. J. (1999) Isolation and characterization of drosocrystallin, a lens crystallin gene of *Drosophila melanogaster*. *Dev.Biol.* 207,204-214.

51. Eigenheer, R. A., Nicolson, S. W., Schegg, K. M., Hull, J. J., and Schooley, D. A. (2002) Identification of a potent antidiuretic factor acting on beetle Malpighian tubules. *Proc. Natl. Acad. Sci. U.S.A* 99,84-89.

52. Baggerman, G., Huybrechts, J., Clynen, E., Hens, K., Harthoorn, L., Van der, H. D., Poulos, C., De Loof, A., and Schoofs, L. (2002) New insights in Adipokinetic Hormone (AKH) precursor processing in *Locusta migratoria* obtained by capillary liquid chromatography – tandem mass spectrometry. *Peptides* 635-644.

53. Pelham, H. R. (1990) The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem.Sci.* 15,483-486.

54. Svensson, M., Skold, K., Svenningsson, P., and Andren, P. E. (2003) Peptidomics-based discovery of novel neuropeptides. *J. Proteome. Res.* 2,213-219.

55. Uhler, M., and Herbert, E. (1983) Complete amino acid sequence of mouse pro-opiomelanocortin derived from the nucleotide sequence of pro-opiomelanocortin cDNA. *J Biol Chem*. 258,257-261.

56. Pan, F.M., and Chang, WC. (1985) Nucleotide sequence of bullfrog pro-opiomelanocortin cDNA. *Nucleic Acids Res*. 17,5843.

57. Martens, GJ., Civelli, O., Herbert, E. (1985) Nucleotide sequence of cloned cDNA for pro-opiomelanocortin in the amphibian *Xenopus laevis*. *J Biol Chem.* 260,13685-13689.

58. Kitahara, N., Nishizawa, T., Gatanaga, T., Okazaki, H., Andoh, T., and Soma, G. (1988) Primary structure of two mRNAs encoding putative salmon alpha-subunits of pituitary glycoprotein hormone. *Comp Biochem Physiol B*. 91,551-556.

59. Salzet, M., Wattez, C., Bulet, P., and Malecha, J. (1994) Isolation and structural characterization of a novel peptide related to gamma-melanocyte stimulating hormone from the brain of the leech *Theromyzon tessulatum*.
FEBS Lett. 348,102-106.

60. Bungart, D., Hilbich, C., Dircksen, H., Keller, R. (1995) Occurrence of analogues of the myotropic neuropeptide orcokinin in the shore crab, *Carcinus maenas*: evidence for a novel neuropeptide family. *Peptides* 16,67-72.

61. Yasuda-Kamatani,Y., and Yasuda, A. (2000) Identification of orcokinin gene-related peptides in the brain of the crayfish *Procambarus clarkii* by the combination of MALDI-TOF and on-line capillary HPLC/Q-Tof mass spectrometries and molecular cloning. *Gen Comp Endocrinol*. 118,161-172.

62. Skiebe, P., Dreger, M., Meseke, M., Evers, J.F., Hucho, F. (2002) Identification of orcokinins in single neurons in the stomatogastric nervous system of the crayfish, *Cherax destructor*. *J Comp Neurol.* 444,245-259.

Legends to the figures.

Table 1: Putative Peptide Precursors: Each putative peptide precursor in *Drosophila melanogaster* encodes at least 2 similar subsequences. The similar subsequences are in bold. The similar amino acids are underlined.

Table 2: Putative Peptide Precursors: Each putative peptide precursor in *Drosophila melanogaster* encodes a similar subsequence to a known peptide. Each putative peptide precursor sequence is accompanied by the known peptide precursor that displays similarities. Both the similar subsequence and the peptide are in bold. The conserved motifs are underlined.

Table 3: Comparison of γ-MSH sequences of different metazoan species to the putative γ-MSH-like sequence in *D. melanogaster*

Table 4: Comparison of Orcokinin sequences of different invertebrate species to the predicted Orcokinin in *D. melanogaster*

Table 5: Comparison of Immune induced peptide sequences (DIM) from *D. melanogaster*

Fig. 1. Construction of database containing all *Drosophila* proteins that are less than 500 residues in length and that have a signal peptide.

Fig. 2. Construction of neuropeptide precursor database across metazoan species.

Fig. 3. Cleavage of protein sequences in the two datasets and BLAST analysis of the obtained subsequences.

Fig. 4. Strategy for the final filtering of the BLAST comparison results.

Fig. 5. Screening method 1. The protein identified by accession number Q9V808 was retained as a putative peptide precursor. By comparing database '*Drosophila* subsequence' with itself, we obtained three similar subsequences in Q9V808 (underlined). The number following the accession number represents a different subsequence within a protein sequence by its position.

Q9V808_6: IPYEVKVDVPQPYIVE
Q9V808_8 : IPYEVKVPVDKPYEVKVPVPQPYEVI
Q9V808_9 : IPYEVKVPVPQPYEVI

Fig. 6. Screening method 2. Q8MS86 was identified as a putative peptide precursor. The subsequence iQ8MS86_2: WKILTAGSHFRWL. is similar to the subsequences P11885_2 (YVMSHFRWNKF) and P06298_8 (NGNYRMHHFRWGSPPKD) derived the corticotropin-lipotropin precursors of *Rana catesbeiana* and *Xenopus laevis* respectively.

Fig. 7: Fragmentation spectra and annotation of 2 peptides identified in the hemolymph a) LDDSENNDQVVGLLDVADQGANHANDGAREA b) LLDVADQGANHANDGAREA Both peptides originate from the CG7734 that was picked up by screening method 1. The most important fragment ions (mainly b an y – type ions are indicated on the spectra)

Fig. 8. Amino acid sequence of CG7734. The peptides that were identified in the hemolymph are indicated in bold. The signal peptide predicted by SignalP is underlined.

Table 1

The known peptide precursors (10)

(1) P09040; Drosulfakinin precursor(DSK) OR CG18090
MPLWALAFCF LVVLPIPAQT TSLQNAKDDR RLQELESKIG
GEIDQPIANL VGPSFSLFGD RRNQKTMSFG RRVPLISRPI
IPIELDLLMD NDDERTKAKR **FDDYGHMRF**G KR**GGDDQFDD**
YGHMRFGR

(2) P10552; FMRFamide precursor(FMRF) OR CG2346
MGIALMFLLA LYQMQSAIHS EIIDTPNYAG NSLQDADSEV
SPSQDNDLVD ALLGNDQTER AELEFRHPIS VIGIDYSKNA
VVLHFQKHGR KPRYKYDPEL EAKRRSVQDN FMHFGKRQAE
QLPPEGSYAG SDELEGMAKR AAMDRYGR**DP KQDFMRF**GR**D**
PKQDFMRFGR **DPKQDFMRF**G R**DPKQDFMRF** GR**DPKQDFMR**
FGR**TPAEDFM RF**GR**TPAEDF MRF**GR**SDNFM RF**GR**SPHEEL**
RSPKQDFMRF GR**PDNFMRF**G RSAPQDFVRS GKMDSNFIRF
GKSLKPAAPE SKPVKSNQGN PGERSPVDKA MTELFKKQEL
QDQQVKNGAQ ATTTQDGSVE QDQFFGQ

(3) Q9VG55; Hugin protein precursor(HUG) OR CG6371
MCGPSYCTLL LIAASCYILV CSHAKSLQGT SKLDLGNHIS
AGSARGSLSP ASPALSEARQ KRAMGDYKEL TDIIDELEEN
SLAQKASATM QVAAMPPQGQ EFDLDTMPPL TYYLLLQKLR
QLQSNGEPAY RVRTPRLGRS IDSWRLLDAE GATGMAGGEE
AIGGQFMQR**M VKKSVPFKPR L**GKRAQVCGG D

(4) Q8SZ21; Tachykinin precursor (TK) OR CG14734
MRPLSGLIAL ALLLLLLLTA PSSAADTETE SSGSPLTPGA
EEPRR**VVKRA PTSSFIGMR**G KKDEEHDTSE GNWLGSGPDP
LDYADEEADS SYAENGRRLK KAPLAFVGLR GKKFIPINNR
LSDVLQSLEE ERLRDSLLQD FFDRVAGRDG SAVGKR**APTG**
FTGMRGKRPA LLAGDDDAEA DEATELQQKR **APVNSFVGMR**
GKKDVSHQHY KRAALSDSYD LRGKQQRFAD FNSKFVAVRG
KKSDLEGNGV GIGEDHEQAL VHPWLYLWGE KR**APNGFLGM**
RGKRPALFE

(5) Q9NIP6; Cardio acceleratory peptide 2b precursor(CAPA OR MT-CAP2B) OR CG15520
MKSMLVHIVL VIFIIAEFST AETDHDKNRR **GANMGLYAFP**
RVGRSDPSLA NSLRDGLEAG VLDGIYGDAS QEDYNEADFQ
KK**ASGLVAFP RV**GRGDAELR KWAHLLALQQ VLDKRTGPSA
SSGLWFGPRL GKRSVDAKSF ADISKGQKEL N

(6) Q9VIQ0; Short neuropeptide F precursor (SNPF) OR CG13968
MFHLKRELSQ GCALALICLV SLQMQQPAQA EVSSAQGEHL
VQPPPEKQSS KDSFLGTPLS NLYDNLLQRE YAGPVVFPNH
QVERK**AQRSP SLRLRF**GRSD PDMLNSIVEK RWFGDVNQKP
IR**SPSLRLRF** GRRDPSLPQM RRTAYDDLLE RELTLNSQQQ
QQQLGTEPDS DLGADYDGLY ERVVRK**PQRL RW**GRSVPQFE
ANNADNEQIE RSQWYNSLLN SDKMRRMLVA LQQQYEIPEN
VASYANDEDT DTDLNNDTSE FQREVRK**PMR LRW**GRSTGKA
PSEQKHTPEE TSSIPPKTQN

(7) Q9VVF7; Allatostatin/MIP precursor(MIP) OR CG6456
MAHTKTRRTY GFLMVLLILG SACGNLVASG SAGSPPSNEP
GGGGLSEQVV LDQLSESDLY GNNKR**AWQSL QSSW**GKRSSS
GDVSDPDIYM TGHFVPLVIT DGTNTIDWDT FERLASGQSA
QQQQQQPLQQ QSQSGEDFDD LAGEPDVEKR **AWKSMNVAW**G
KRRQAQGWNK FRGAWGKR**EP TWNNLKGMW**G KR**DQWQKLHG**
GWGKRSQLPS N

(8) Q9W0W6; Neuropeptide like 1 precursor (NPLP1) OR CG3441
MQAVLQSAHS SRRLMLLLSM LLNAAIQPRS IIVSATDDVA
NVSPCEMESL INQLMSPSPE YQLHASALRN QLKNLLRERQ
LAVGEEQPLG EYPDYLEEDK RSVAALAAQG LLNAPKR**SLA**
TLAKNGQLPT AEPGEDYGDA DSGEPSEQKR YIGSLARAGG
LMTYGKR**NVG TLARDFQLPI PN**GKRNIATM ARLQSAPSTH
RDPKRNVAAV ARYNSQHGHI QRAGAEKRNL GALKSSPVHG
VQQKREDEEM LLPAAAPDYA DPMQSYWWYP SYAGYADLDW
NDYRRAEKRF LGRVLPPTRA TASTHRSRL

(9) Q9VC44; Allatostatin precursor(AST) OR BCDNA:RE16553 OR CG13633
MNSLHAHLLL LAVCCVGYIA SSPVIGQDQR SGDSDADVLL
AADEMADNGG DNIDKR**VERY AFGL**GRRAYM YTNGGPGMKR
LPVYNFGLGK R**SRPYSFGL**G KRSDYDYDQD NEIDYRVPPA
NYLAAERAVR PGRQNKR**TTR PQPFNFGL**GR R

(10) Q8ML70; Immune induced protein 10 (IM10) or CG18279.
MKSFGLIALA ICGVICVAAE PQHTYDGRNG PHVFGSPGNQ
VYIRGQNEGT YSVPGVGGQF QNAPQRGEHV YTDEAGNTFV
NRKNAGGPAS HTISGPNFSA KNLGPNGAKS VGIPQRARR**S**
PQFHVERPGR TVDVGNGGFY IQRGRR**SPQL HVARPDRTVT**
IGNGGVYIQR SRR**SPQFHVE RPDRTVDFGN GGFSAQRF**RR
GINDARVQGE NFVARDDQAG IWDNNVSVWK RPDGRTVTID
RNGHTIVSGR GRPAQHY

The putative peptide precursors(47)

(1) Q8IP68; CG31813.
MSAKYTLIFA LAALCCLVFS TEAAAQRSRV LSSRRGSELV
EKTSDNKEDS ELAAQEQDLE RQEQEEQNDR LEGRSDDVAE
GSDNKEDKET ATNNKDTIVK PNKDDARARR IVRAGRRRGG
RRGGRRGGRR SARK**SVRRGG** RR**GGRRRGG**R RGRGGARRRT
SVKRRSGKGN KA

(2) Q8IPI9; CG32829.
MHGHFKLLLV LGLLLTYAAA KKKEASSSEE EEGGDKYEKK
KFGDEEHKGE HGHKEHKEWE EDEKKHHEVE DHEHHHGDKG
SKKKK**HYDEK DEHGEKHEHG AH**KKGGKHHH KKKHKKGHKE
LEYHKKFKKD EYIKEKKFYD DEHKGGHHKK YGKEHHHHAE
EHGEHKK**GEK HEG**GKKKGHK KHHGHYKKGH HDEDHKKYKK
EHKYGDSFEE KKEHGEKGSK KHGHKHYKKK GGKH

(3) Q8MS63; BCDNA:LP08232
MRVITVNLLL MAICACSTLW PANALQCYSC VGNECHVETV
PTVTCTLDDV SAAFGLDTNA AFNPFKRSLL SVMPRAADDD
STTGSTTIED SSTTASASTT DSSTASSTTI GESSSSSLGS
STSDSSTSDS TTDSSTASST TIGDSSSSSL GSSTSDSSTS
DSTTDSSTAS STTIGDSSTS SLGSSTSDSS TSDSTTDSST
ASSTTIGDSS TSSLWALLRQ IPQLQILQQI RR**LLPPPQLE**
ILAAALWVLP RQILQPTILQ QIRR**LLPPPQ LEILAAALWV**
LPRQILQPPI LQQIRR**LLPP PQLEILAAAH WVPQRQIL**RL
FPPPQMQVQP PVLQLLRHP

(4) Q8SZG8; LP03261p or CG11470.
MFAKQAKILV FCLSLGLSLA **VKLQIPFRPD AHIQELQLQS**
RELAEVHSDH STQCFSIYKP KLAKIADQFE TNFTACISAY
DNCTSHISEK YAEDRQILLR SANIGCSYPN SNCQVWTLEQ
QPLDTVVSRL ECASTNSAES SKTFYAISAN ATQIAVQIQE
QYTILESRK**S VCINDANRSR LECASTNSAE SSKTFYAISA**
NATQIAVQIQ EQYTILESRK SVCINDANRS YVEDTSDTYE
LLNNCLKNGP TTTTCNPLTY PTTTTATTTT TTITTAAPLL
R

(5) Q960I1; LP07079p or CG12164.
MRSLSLLLLS ASCALIISLA YAHPPEGVWK KKLTWKEDWV
QVWKTVKKEA WETKWKKVSV PIWKEVKVPV WKEEKVPDWK
IVKK**PKIEER EVPAWKEVKV AEW**KKITKPI WVPTKVAVWK
EIQVPIWKEV QVPYWKEIQV PIWKEVQVAD WKQMFEPQWV
KMGIPGEKFL GKDHEGWEYT SHDLWRKKLI WKPVWKKVWR
TDKKQEWKTE KKQEWRTEKK **QEWKTEKVQE WKQD**KK**LEWK**
DEWIQVWKPV KKQIWIKEKR **ETWIEEKVQI WRTE**KRQVWA
TEKR**QAWKDE WQSVNVPVWK EVKVQEW**KKV WKPVWEKVWV
PVSHGHGWD

(6) Q9NEG3;EG:BACR43E12.5 OR CG14418
MCGKGFGMIL VLIWATALSS DLSLVSGQTV APLTSSTSAP
PKMEVKPTKP MIVTGFITKS GNIYEIEDKR GSIGSIEGRQ
ADQEQIVCNY GNVVIYSDVP CDQVKNVRVG EVKPLKDDRV
EVTTEKNHAD GEGEQQSQDM QQEQQQQDAA EDQDQEQDQD
HQSVQSNHPR GQQNNRRRRR R**PQQQQQQQR LQQQ**RRRR**QQ**
QQKLQQRRRN GNGNGNNSLR RRR**NRNNNNN NL**KRQQQRRR
R**PGNNNNNN**R R**RLNNNNSNN NNKQRQQQH**R R**QQQQQQ**RRR
R**PNRNINRQQ QQQ**RRRLHDD NI

(11) Q9V401; EG:96G10.8 OR CG14265.
MKLHWLLLAV VLICALYSAT GTSPTTETST STESTTATGS
STSTTSASTS SSSSDTTEAS SSSSDSSTSS SSSSSSSSSS
SKKK**AAARR**R R**AARRRRRAA** RRRRAAAARR RAQRRRNRG
---
(12) Q9V5U4; CG13227.
MDHKWIIFFF SIAALLLCNF VRADETETEV VQEKPSSIQL
LDAGETAQSD STDENVRK**VR QYFGPPPFGP PPPPFFGPPP**
PPYYGGGFGG GFGGGFQRTR VVTRTRYRGR **GGYYGGGFYG**

(13) Q9V7U4; CG15615.
MICHQLAHLT QLGLLGVLCL AIGGSQSKAV VDQQPAASQV
DSKSVAQQRI DLGLGLGDDL IHDHHHEHII EHHEHHHEHH
DPGYWKKKVT WKEGWKKIWN PAKKQIWNPS WKKIWKPHWV
KVPGWKEIQV PAWKQIWVPH WKEILVPAWK DIQVPDYKQI
WTPELVKVGI PGEKYLGKDH EGWEYTSHDL WKKKVVWKSH
WKKIWKPAKK QIWVPEKK**LE WKEAWKQYWK PA**KK**EIWTDK**
LEWKEAWKQI WVPGWKEIWV PGWKKIWKPV VISEWFPSPD
HHDHHHHEHH DWDRKDTGVT ATRTADGKDK VVWKRDDTNA
AGKPTMLQPV ASADFQAKSL EAAKSPVAAP AASVATTSQS
FKFPGA

(14) Q9V808; CG30101 OR CG6564 OR CG15901.
MRMFVLPCLA VCVALAHCGG AVEDKKAEGD GKTVEKRGLH
LGDYHHYQPH HEHIKTVTIE KKIPVPYTVT KHVPYTVEKK
IPYEVKVDVP QPYIVEKKVP VHVKEYVKVP VHVPKPYEVI
KK**IPYEVKVP VDKPYEVKVP VPQPYEVI**KK **IPYEVKVPVP**
QPYEVIKKVP HEVKVEVPVP KPYEVIKKVP YEVKYEVEKP
YDVEVPKPYD VEVEKPYTVV VEKKVPYEVK VPVDKPYKVE
VEKPYPVHVK VPVPQPYTVE KKVPYTVEKP VPYEVKVPIE
KPIPVYTEVK VPIHKEIPVP EKYHVEVPIF KHHQEDHHDY
HSHGHGHY

(15) Q9VDD1; CG5862.
MDLIILVGIA VALLVVIVTL YLLQKKNAAP ETKVAAAPQR
GVPQRAQEGV PRRAQIARNQ RNRLRQNVPA APVAAAAGAL
PAAGDSDHED EGQVDGDEAR VPQGAVLDEK MGAKKRAKME
AKEQKRLQRE QELHDREQRK **VKEAKEEAER K**QQEDLEAEA
ERK**RVDAERL AKEERERK**EH EEYLKMKAAF SVEEEGFEEG
DADDQDNLLA DFIQYIRDNK VVVLEDLAVA FKLKTQQVID
RIQNLQADGT LTGVIDDRGK FIYVSEKELL AVAKFIKQRG
RVSIAELAES SNNLINLTPI SAGGGEASS

(16) Q9VFI4; CG14852.
MRTTTLLLSL GLLVLCFSSY SFAEDDPTDG STTPTDGSTT
PTDGSTTPTD GSTTPTDGSN TPTDGSTTPT DGSTTPTDGS
TTPTDGSTTP TDGSTSPSTS PSTGDNTSPS TGSPDSTPDS
GSGSSNNSGN NKRNNRRRRR **QRAQRA**RRR**R AQQA**RRRRNR
RNNRNNRLSR TNSI

(17) Q9VFI7; CG14850.
MRATSIILSG VLVLVACLLR SSEAVTCTAD PNVTGCIDCT
TSPSDPECVA EAANTTTKPA DGTDTTTPTT GGSTDATPAG
STTPTSPSGT VTPAPTSSPS DSTSPSDSTP TSNNAAAVAR
RRRRMAARR**R AQRRRRAQ**RR **RDQRRRRAQ**R RRRQNQSG

(18) Q9VTF1; CG32071 OR CG6261.
MASRIRREVT MRPILVLSLV ILATLVVLSS QATSTSPTSS
SSTSPTSSSS TSPTSSSSST SSATTTTTTT TTTTAATTTS
TTTEKSKKKR NRRRRRRIII RRRERR**DERG ERS**RR**ERGDE**
RGERSRRNRG SDGERIVRRY VVRERRFRRY

(19) Q9VUE4; CG3868.
MKLFAICALL ILPLVSG**GIV PRTPTRHFVA YEIHSTHNLP**
QMEAIVQKMQ ILFQNMIPDM SVLADAVAVP DMSVQSDAGV
LPDMSVATDA DETPRR**RSAE LKVAPIAKIA APMIAKIPDM**
SVQADAGLIP DISVASDADE TPKRASFGLK K**PLLDRIAGA**
IAKMDADLIP DMSVATDGAD TPKQAVSGDK ALPKMTKVSG
GEPVEIKTHA DADQIPDMSV ASDADDTPRK VAASLKTLPK
MTKVSGGKPV EIMTHADDGL IPDMSVASDA DVTPKQTVAG
LKALPQLTKV SGANPVEPIS QANADLVPDV SVASDAEDKA
ANQLPDISVA GDAKFDPKMT LAALLAT

(20) Q9VX67; CG5172
MRHAVILVFV CCLLIALTSA GLLGGGGGGG GYGGGGGGGY
GGGGGGQSGY GGGGQKNGGG GHGGGGQGSY GGGSQGK**WTW**
RRWPRRLAEE RRRWWTRR**LW RWQSRWP**RRT RRLWRRRSWR
RWPCKQVLGR QSR

(21) Q9VZX8; CG12093.
MFIWHYLTLL CLGFALIDAK SIKKSQANLL EPTLDPDEKL
GKIGKSPPDI SNDVQRVQVP LGATPGKNGW QGK**WFPHAPG**
QQQEAIELKM KSTTEEVPPL EFTTKDGWQG K**WFPQAPGEP**
HLVKKKPLVD NKTKSGKSTT QDTWQGKLFP QGPNESHKTK
QSSCDDGKSN LAVDFDPNDI RDSLNFLCIS SNHSLYQPDL
SREAILTEHF LPSAYLPPAK CLNESISYSH LPATNGPYRP
MPAEYGTYSY LPPQRYVRNL AEGAIVMLYH PCAFPGQVKQ
LQDIVGGCLY RHLVSPSLAL SPQRPLALLA WSRSLEMSVV
DRQLAADFIQ KHAKQGPLAP EELSRLIVKR QTYKEGLLRE
AHLVNTADDY ELCGYLQEDM

(22) Q9I7M6; CG5559
MDIVIREEDI SLAQIGVYAS VSFLVVSAVG AALYTTCSKR
YRLNWFEQNL LESANEKDED QQREALVAGA VGYNVDNVNE
VPRGKYSSGN AGNLSPTSLK SEDNDPAFWV PASVTSTAAI
QQQVSNTTEE SAPPTSPTGS LKSNTLSYCS TTSVPIARSD
KHVVLAMHPS RPRVSSMNAK LDHTKIDMTL YRSHSQPKTI
NPVSLNEVRG NLHVSLGYDP VGGLLNVRLL EAQNLQPRQF
SGTADPYAKV RLLPDKKNFW QTRIHKR**TLN PVFDEQFVFE**
VTAGVIDKRT VEILLYDFDA YSRHVCIGGS KLHLANLDLS
EQLKLWTPLS SASAQDMKVD LGDIMVSLAY LPSAERLMVV
LIKARNLRIV DDARNSSDPY VKVTLLGPGG KKIKKRK**TGV**
QRGTLNPVYN EALAFDVAKE TLKNCVLEFT VVHDGLLGSS
EILGRTLIGN SPEVRTEEKI FFEEVFRAKN ATAQWVPLQE
PANNLATSAK SSKN

(23) Q9VTC3; CG6409.
MKSVIILLAL VAFCHAAPLD VKESTSEELS RPSPISPDVL
VDPKPTAKVV LLKDAPVLNR QRRNEPKKPD SVEAHEQLHH
SQPAVPIHDQ KPSPAGPQDH KRKR**EAHHEE GHHGDDAVPQ**
KEEVKQAPED HDAKKEKR**EA HHEEGHQDHE APHKEEVKQA**
PEAHDAKREK **REAHHEEGHH GDDAAPHKEE VKQAPETHED**
KKEKREAHHE EGHKDEAGDR PQVEDLSLPH ALPAVAHTAE
LPKKQEKRET VEKPDSVKAR ESLQHNPQRS QELHKAIESL
AGGKAPEQRH QRDIPVPTQT KATTTTDLPS TTKSELESTT
PSIHNLHIPH PIPVAELFEK NKHADKSTSS SEESKEKAKA

(24) BK002432; HDC12790.
MRLETNGIRL WSYVVINLLL SDSNAL**FKFK NVKCTCYEKS**
FCELKR**CELK VL**GRGIVGLN LHAQVYKLPI KSTTHDIIVN
QMVLNDDMIS KAPVPNGFYK LRFIVKTDGV WRGEVEVHAE
VNLGIDR

(25) BK002548; HDC13589.
MKLSLVLFVL SMVLYVAHVR **AADSSTSTES STSTTDSTTT**
ESTTESTTSS SSSSSSGSNK KIVRLSNLKY SITRK**IRVGS**
TTSSSTRSKS SRAKSRKARL NAAKRSKKAA NRKLNKKSKN
RNVRVVRG

(26)BK002769; HDC15078.
MQLSIIVLLL CSVVVA**NSLA PSQLASKTPS PNVSKDQPSS**
REEKPSLKPC KPIETVSSEP KGLGNTPKVG SITPESAKTS
GTTVDKSLDD CEPIPEGIGS RLNARTLQTL PSSKQVSHIK
QCKTVETTSL EPNEVLQSTI SPEVSSTRKT VPLPTLAPSS
SLTSTEINTD IDSISRKSER RVVPLPTLKP TSTLSPIQVG
SDYGQR**NVPH SGIMGLIPLA TRALELIKER VDESVPNPQK**
LLPQPTLKPS STLTTTEVSL EREDQIQDDC ETLPDGIGSR
LDAKTLKTLV KNQLG

(27) BK003517; HDC02932.
MAQLFSICLL IATVVNVHGF SKYGRDCRDI LCAPGQKCII
SRDPCSGSNK LENTQCGKYP TCVMHNHYSS TSGETLERGK
RQANQQGYGM GGSGMGR**PNG MNNGGNGGNG MGGQYGNGMG**
GQNRNGMGGQ YGNGMGGQNG NGMGGNGMRG NGMGPGGMGG
R**GGNGNGMGQ GGMGGNGMGS GGMGGNGMGG NGMGGNGMGG**
NGMGGNGMGP GGMGGNGMGG NGLGPGGMGG NGMGPGGMGG
NGMGGQGGYN GRWGQNGMGG PNGMGGRNGM GRPNGMGGPP
GGQNGMGGPP GGPNGMGPGN WQGNNNWGNG NNGNGSNRYG
GNSYSTTPTT TTDGW

(28) Q9NKE1;BG:DS00180.7 OR CG16882 OR CG31839
MRSSSCQLVT LGVLLAICSL GQGQFKTAGI KTRQPPSGNL
QLAGNSSESG WRSYNQSSYG WSTQNQSNYA WNQQNHVEQG
SAGFVRAEVF QPVTLPPLYG HYVQPVTPPA HRVQVLDETA
LFINKTRSAM ASGVCYKEVP TASLLRNSRD QFVGNGTTPD
MSRIQVCCDG YERNPHIYRR **CEPICADDCR NGICTAPNTC**
VCIPGHVRTA EGKCISTCPL GCGNGVCDER NECKCREGYS
LEPETRKYCQ PECKPGCSFG RCVAPNKCAC LDGYRLAADG
SCEPVCDSCE NGKCTAPGHC NCNAGYLKLQ GR**CEPICSIP**
CKNGRCIGPD ICECASGFEW DRKSAECLPK CDLPCLNGVC
VGNNQCDCKT GYVRDEHQRN ICQPHCPQGC QNGYCSAPNF
CICRPGFIKS GIKGRQTCQA V

(29) Q9VNP6; CG11131.
MKLFTAVLAI CLVAFAAAQS ADPAAALEPS SEYLPPVGEA
EAAQLSENGY KYRTVRRLKL RHRR**EVPNQE YLPPVENAPS**
QEYLPPVDAA AIGDTKVADD GYRYKTVRKL KFRARHRR**DV**
SEIAEPSGEY LPPVQVELAP ELKTILGDDG YKYKTVRRLK
FRRHRR**EAVA EEAAAESAPN GEYLPPAEAA AAAPAAAEAE**
PKSAEEGTEL AKDGYRYKTV RRLRYRYRH

(30) Q9VZB4; CG7465.
MKLFLVAFAV IAAVAADVSH LPSNEYLPPV QEQQIIAGPS
NEYLPPVQAE SAPAHELADD GYRYKTHKRV VVRRHRR**DVN**
ELFNEYLPPF AAPSNEYLAP AEGAPETILA DDGYRYKTHK
RVVTRRHRRD VSHLPSNEYL PPVQAAAPSN EYLPPVSAPV
QVAAPAPAPV QIAAPVQLAA PAPVVVEAEP AHELADDGYR
YKTHRRVVYR RHRR**DVNELS NEYLPPFAAP SNEYLAPAET**
APETDLAVDG YRYKTHKRVV TRRHRR**DVNE LSNEYLPPVQ**
SAPSAEYLAP QENTVEAAPA HVLADDGYVY KTHKRVVLRR
H

(31) Q9Y0V9; MIPLE OR CG1221.
MRINCNALFL ASLVTWSGVM CSTVLGTTEG QETPLALPVA
EQTQPTTAIQ GEVWEEDDHE VLIRNERGTK SDGLSCRYGK
NPWTECDTKT NTRSRTLTLK K**GDPACDQTR TIQ**KKCKK**AC**
RYEKGSWSEC ATGQMTRADK LKASSDPSCE ATRVIKKNCK
PGKSKDKSAK EQRKNIDKAA RKGRV

(32) Q9V5V1; CG7738.
MRLTLLALIG VLCLACAYA**L DDSENNDQVV GLLDVADQGA**
NHANDGAREA RQLGGWGGGW GGR**GGWG**GR**G GWG**GR**GGWG**G
R**GGWGGGWG**G R**GGWG**GRGGG WYGR

(33) Q9VRI3; CG10918.
MRAYIAITLL ALVAVVVAQG GGGRRGGRGG GGGGGRSLGG
FGGRGGGGFG GR**GGPGGTGG PGGFGGP**GR**F GGPGGLGGGG**
GFGGPGRFGG PGSFNGGFGG PGGWGPRPRP RRPRPWWTTT
ESSLADTSST TSSSSSSTDS SSSASSTDSS TDSSTDSSTD
SSTDSSTDSS TDSSSTESTT GSG

(34) Q9W2R5; CG15225
MPSVNQFKAN TLFTMSTKLV LFLCLALFGA MAMAIPLEEN
LEQNSEVENS QELNSVEEND AEADPSTWRK **LFGSEGSAGN**
LVVSFPQTTT TTTKK**SKTTK RPYYPYPPKY PPYPPYPHYS**
PYPAPPYPYP GYYPPPPPPY PYPYPGHGGH SGSGGHDSGH
GGHHHHTTTT TTTTTTTKKP NHGQYPPPPP PPPYYPPYPY
YPPPPPPPPL PPPSNDSGES SETSAKICKK FFGIGLICT

(35) P26023; Pre-intermoult gene-1 protein precursor (PIG1 OR PIG-1 OR GSG) OR CG10790.
MKLTKLWLLF VCLGLFVTLV VSADTDSDAD SDSSADSDEN
TTASGSIVTS TTESSATNSS GSSDDASGSS SDVDDGSDDD
TDSGSDTDYD TPTTAPVVKK RANRKKANNN KKRASNNRK**K**
ANNNNNNNKK **RANSNNN**RKR R**ASNNNN**KK**K ASNNNSN**RRR
NNNSRRRG

(36) Q27320;EIG71EE OR CG7604.
MKLTVVCLVV SFFLLHYAEH SDACLEVIEK ALGLQPCNEG
GRNEHREPHR GGPGPVRSTR RRGRIPRRRE TPRPIHHNTR
ERRHHTKTRK PRKPVPCITK RTEPPPVTDF TTRK**SNPPCT**
CTESTTRK**TN PTCTCTESTT** KK**TNPTCTCT ESTT**KK**TNPT**
CTCTESTTPL TEPPVTDITT QKSNPPCTCT ESTTRK**TNPT**
CTCTESTTQK TNPTCTCTES TTKK**TNPTCT CTESTTPLTE**
PPVTDITTQK SNPPCTCTES TTRK**TNPTCT CTESTT**RK**TN**
PTCTCTESTT KKTNPTCTCT ESTTPLTEPP VTDITTQKSN
PPCTCTESTT QKIKSTSTTQ GTEPPSTQKT LPPNPPSTKN
TEPPNSTPPP EKTTRKPCGC SSSHPSGWNL AVL

(37) Q9VUS6; EIG71EC protein(EIG71EC OR CG7608)
MSKITLIFAI LCLCVAVQAQ TREQEICRQE NETCRRNERR
LGVQNDVSTT FNNHCRRQSG IRNWRNVSRC ELSLATCRLT
LERCAVINCK NVRNSIDGGV TARPPTSRRT TRRIPDTRRP
RTTRRTPTTR RAPTTRSSRR TTRR**RAPTTR TT**RRTTRR**RA**
PTTRTTTRRP TEE

(38) Q9VV46; CG4784
MHCFTWTILG GLLALTSAAG LPQRPSSGYQ EQDTARAFYS
YGYRDENAAR AEYSSRDGTS RGFYSYVDAD GKLQTVRYEA
NGVQGFKAEA SNQPQAPVDK GKAPLPVTDT EEVQQARLNH
LNALREAREK ALATSLREEA DRRQQEQIRN NNEDQQSGEQ
SLTDEDAAIL ERVRAELSAM LADRQRELNL PRNRDDREQR
EKQEIRQDQR K**ELRQDLRQE QRQDQRED**RR **QDQRED**RR**QN**
QREDRR**QDQR EE**RR**EDQREE RGEDQREE**RR EDRREDRRED
RRQDLRIDQE SRQDLLQDLR QDLRQDLRQE LRQDLRQDQS
RNQESLRDSS QIRENARQIS SDRDGDDLRL RTVYSLADLS
SSSYLKLGDL ASAEKLLEDR LDNSDLRVPI GAYYTLVSPN
TKYSVTTPTE LRTLRPVALS RSLLVSKRN

(39) O61351; LCS protein(LCS OR CG12794).
MRTLILVTLV ALVAVASA**QG PGPW**GPGGPG GPGGPGRGRG
GPGR**GPGGPG GPG**GR**GPGGP GGPGGPGGPG GPGGPGGPGG**
PGCPGGPGGP GGPKPWGPPS NQTTSTTTEA STSTSTTTAS
STTVSSTTES STESSTESST ASSTE

(40) Q9VV44; CG4818.
MKYPLLLLGS LSLAHGLALY YPYAYTAEGS AVFTPTQRQY
IAKDELGQYS YGYSEPLSSK QETRTLDGIT QGYYSYRDAA
GK**LQTVNYVA DNKGFHVAAT NLPKAKVPQE SLEFSPRSAS**
HPVDHHVEHH AEVSHAVVQH PVGHHPIEVP HHHTVVESGR
SAHPDGHHPV EHHEHRVAVA QHPVGHHPVE VPHHHTVVET
GR**SAHPDGHH PVEHHEHPVA VAQHPVGNHP VEVPHHHTVV**
ESGRSAHPEV PHSIEHHEHP VSGSDPSGSH GGHSQLPHPV
SDTAEVAAAK SLHLQRVHDE GVRNQVLAKI PVAVARSHHV
VVPAPVYGYT IPRYYTPGFY Y

(41)Q9V6U2;CG6357.
MTSARILLGV PLLLYLMGVA LGVPVSTSSP ATQKINPEIG
VTTGKSDADS STPTIEHTSG LSEFEEECQF AWQRFLVDFD
VHYDNDYERQ KRRDIFCENW QKVRDHNLKY DLGVVSFKKG
INQWSDLTFE EWKEKQTPKV MPEIASESSK EERDKVNCQA
AWEKFLIDFG AQYKNANETE KRRNVFCANW RAIVEHNVQY
EKWAEPFKRD INQWTDHTIE ERSSPAPEIR KEEATTSTSE
IDNDNIICQP AWEKFLIDFK PSYQDDTETE KRRNVFCDNF
KSIHKHNVQF DLGNISFKK**G INQWSDLTVE EWKNKQRPAF**
NPEFSKVEAT TKISKDKRDD NTCQAAWKKF LIDFGAKYQD
EKETEKRRTI FCDNWKAIQE HNEQFELGVE SFKK**GINQWS**
DLTVEEWKTK QRPNLAPEFS KEETTTKISK EKKYKKNVF

(42) Q9VRU9; CG12330.
MRLTTLFSLI CIAIGYVRS**Q PAGYPSARPP ATYLPVKPPA**
PPPRPPPPAP ANSYGPPKKG NGK**PPPAPPK PSYGPPPKNG**
NGKPPPSNAY LPPGNGNGGS SGGGGAGGGG GEDIPIIKLE
SKVNTDGSYM YEYETGNGIK AEEMGYLKNA GVEGAEAQTA
EGSFSYTSPE GQEISLTYIA DENGFQPQGD HLPTPPPIPI
EIQEALDKLA AGGGCHGCDD NETGGNDDGG GGGYVYRRK

(43) Q9VKE2;CRY OR CG16963.
MKRTYLLLCL SLLTCNVANS AYLRPIDLNQ LAKSSNLQQQ
QQQQLRGALN RDDNNDDDDA TTLAPNSNED YDTRPQYSFA
YDVRDSLTGD DKRQEEKRDG DLVKGQYSLI EPDGTRRIVE
YTADDVSGFN AIVSKQRLDE QQQQRLSAST SSRFNSLEEL
QTRLTAQAIA EAQSLVEAQQ ASQLQLEAQN RRESENQARN
QAQQLMEQFQ QQVQQQEQQR LQQEQQLR**DL QRLQEQRDRE**
ERDREQRERE QREREQRERQ QREQELRERE LRERELRD**RE**
LRDRELRDRE QRDREQRDRE DRRQQAERRQ SQNQRLLDQQ
TLLLAQSLPS IQATVVSHPP TLLATRLPVS TTTSSSRTTT
LLTRERDLEA WRQLPNARIT IDRSSQPLIL SQPSSATVQA
QLISSPLLLE SGNLNGLITT RLNGNAARAG AALSWSNGRR
LLNNDLWQLD RLDDRADNRR ESEELRSNSA ERRSKNW

(44) Q960H8; LP07813p OR CG8502
MFTKSMLSFS LVVALFVVCH ASPVPDNNGR SGSQESIGR**Y**
HHMPIPYRHV SDQRELGK**YH HIPYPYDGGY GPYAGSNIPY**
VHDDRPYNHD LYTSTTTKKP TTTTKRTTTS TTTTTTTPRN
ILFNYDDEGR HKILHKEEVR KQDKYDHSYL TENGIYGEEQ
AKLHHTGGTH AKGFYEYTGD DGKLYRVNYA SNDGGFMPQG
DHIHPIPDAI VRALKYVEEQ HKINGGAQFD HRGFRINHMT
KDLKAQIKAI HLEEMPKELT EQIHMLEHEV ELAEEEEREE
QAALERLRQA AKSH

(45) Q9VWG7; CG14218.
MLRLQLMALL LVGLLLALIS ADPGPQDTEV AREKRGTITL
DFGLLLRNLL LKSAQLSSAK ANLARTTRR**P PTTTTTTPPP**
PPPPAPIRIR K**PIWHPFFSS GFLPGAFDVD YADPPAPRPP**
APAPPTTQPP RRVRPQVRPR PRPTTLAPPP PPNYDYDYDY
DAQPAAPAEP PPPPPPPPPP TAPPRPRPRP RPRPQQPDPQ
QRRPAQLGDR LIYQYAQPTD TFFRSRAVAE ATAAADPDSG
DLVD

(46) Q9W4G9; CG2871.
MWQIFDILLT LFYLFGPLVI FFAYIFYSLS LMRADSHELD
LYNENRRRR**E SIGGDDQGSP RATSERQSRE EAREELEIGC**
NESNSSEFEW YDEDARRRGS FGEDDLGCPR ATSKR**QSREA**
AREELEIGCN ESSSREFEWY DEDARRRGSF GGDGLDSPRA
TSKR**QSREAA REELEIGCNE SSSREFEWYD EDA**RRRGSFG
GDGLDSPRAT SKR**QSREAAR EELEIGCNES SSREFEWYDE**
DARRRGSFGG NGLDSPRATS KR**QSREAARE ELEIGCNESS**
SREFEWYDED ARRRGSFGGD GPDSPRATSK R**QSREEAGEE**
LEIGCNESSS REFEWYDEDA RRRGSFGEDD LGCPRATSKR
QSREEAGEEL QIGCDESNSV **CSRGNPENSD ATRYYYVNSG**
PHSKMFSK

(47) Q9VP98; CG11458.
MRFALLAVLL IGVIFAFVSA **GGGGGGGGGQ GGWQKNGGGG**
GGGGQGGWQK GGGGGGGGK**H GGGGGGG**GK**H GGGNGGG**GK**H**
GGGGGGGGGG GKHGGGGW

Table 2

The known peptide precursors (42)

(1) O96690; Pigment-dispersing factor (PDF) OR BCDNA:RH08487 OR CG6496.
MARFTYLVAL VLLAICCQWG YCGAMAMPDE ERYVRKEYNR DLLDWFNNVG VGQFSPGQVA TLCRYPLILE NSLGPSVPIR K**RNSELINSL LSLPKNMNDA** GK

Q9TWW7; Pigment-dispersing hormone. Procambarus clarkii.
NSELINSILG LPKVMNEA

(2) P09040; Drosulfakinin (DSK) OR CG18090
MPLWALAFCF LVVLPIPAQT TSLQNAKDDR RLQELESKIG GEIDQPIANL VGPSFSLFGD RRNQKTMSFG RRVPLISRPI IPIELDLLMD NDDERTKAKR **FDDYGHMRF**G KR**GGDDQFDD YGHMRF**GR

P41492; Neosulfakinin-I.Sarcophaga bullata.
FDDYGHMRF

(3) P10552; FMRFamide (FMRF) OR CG2346
MGIALMFLLA LYQMQSAIHS EIIDTPNYAG NSLQDADSEV SPSQDNDLVD ALLGNDQTER AELEFRHPIS VIGIDYSKNA VVLHFQKHGR KPRYKYDPEL EAKRR**SVQDN FMHF**GKRQAE QLPPEGSYAG SDELEGMAKR AAMDRYGR**DP KQDFMRF**GR**D PKQDFMRF**GR **DPKQDFMRF**G R**DPKQDFMRF** GR**DPKQDFMR F**GRT**PAEDFM RF**GRT**PAEDF MRF**GR**SDNFM RF**GR**SPHEEL RSPKQDFMRF** GR**PDNFMRF**G RSAPQDFVRS GKMDSNFIRF GKSLKPAAPE SKPVKSNQGN PGERSPVDKA MTELFKKQEL QDQQVKNGAQ ATTTQDGSVE QDQFFGQ

P01162; FMRFamide. Helisoma trivolvis.
FMRF

(4) P17975; Adipokinetic hormone precursor (AKH) OR CG1171
MNPKSEVLIA AVLFMLLACV QC**QLTFSPDW** GKRSVGGAGP GTFFETQQGN CKTSNEMLLE IFRFVQSQAQ LFLDCKHRE

P08379; Adipokinetic hormone II precursor. Locusta migratoria.
MTQSCTLTLV LVVAVLAALA TA**QLNFSAGW** GRRYADPNAD PMAFLYRLIQ IEARKLAGCS D

(5) P41494; Dromyosuppressin precursor(DMS OR NEMS) OR CG6440
MSFAQFFVAC CLAIVLLAVS NTRAAVQGPP LCQSGIVEEM PPHIRKVCQA LENSDQLTSA LKSYINNEAS ALVANSDDLL KNYNKR**TDVD HVFLRF**GKRR

P41855; FMRFamide-like neuropeptides precursor(FLP-1 OR F23B2.5). Caenorhabditis elegans.
MTLLYQVGLL LLVAATYKVS AECCTPGATS DFCTVFSMLS TMEQNEVMNF IGENCDGDAE VALQKMEKRK PNFMRYGRSA AVKSLGKK**AG SDPNFLRF**GR **SQPNFLRF**GK **ASGDPNFLRF** GR**SDPNFLRF** GK**AAADPNFL RF**GKR**SADPN FLRF**GRSFDN FDRESRK**PNF LRF**GK

(6) P81829; Leucokinin precursor(LK) OR PP OR CG13480
MVLLAFGRQV YGASLVPAPI SEQDPELATC ELQLSKYRRF ILQAILSFED VCDAYSSRPG GQDSDSEGWP FRHYAPPPTS QRGEIWAFFR LLMAQFGDKE FSPIIRDAVI ERCRIKSQLQ RDEKRNSVVL GKK**QRFHSWG** GKRSPEPPIL PDY

P21143; Leucokinin IV.Leucophaea maderae.
DASFHSWG

(7) Q07892; Eclosion hormone precursor (EH) OR CG5400
MNCKPLILCT FVAVAMCLVH FGNALPAISH YTHKRFDSMG GIDFVQVCLN NCVQCKTMLG DYFQGQTCAL SCLKFKGK**AI PDCEDIASIA PFLNALE**

P25331;Eclosion hormone precursor.Bombyx mori.
MANKLTAVIV VALAVAFMVN LDYANCSPAI ASSYDAMEIC IENCAQCKKM FGPWFEGSLC AESCIKARGK **DIPECESFAS ISPFLNKL**

(8) Q24049; Amnesiac neuropeptide precursor (AMN)
MLWRCTAYYC FTLFFLLFRA SALRRRVVSG SKGSAALALC RQFEQLSASR RERAEECRTT QLRYHYHRNG AQ**S**RSLCAAV LCCKRSYIPR PNFSCFSLVF PVGQRFAAAR TRFGPTLVAS WPLCNDSETK VLTKWPSCSL IGRR**SVPRGQ PKFSRENPRA LSPSLLGEMR**

P80090; Molluscan insulin-related peptide 3 precursor. Lymnaea stagnalis.
MASVHLTLTK AFMVTVFLTL LLNVSITR**GT TQHTCSILSR PHPRGLCGST LANMVQWLCS TYTTSSKV**KR QAEPDEEDDA MSKIMISKKR ALSYLTKRES RPSIVCECCF NQCTVQELLA YC

(9)Q26377; Crz precursor(CRZ) OR CG3302.
MLRLLLLPLF LFTLSMCMG**Q TFQYSRGWTN** GKRSFNAASP LLANGHLHRA SELGLTDLYD LQDWSSDRRL ERCLSQLQRS LIARNCVPGS DFNANRVDPD PENSAHPRLS NSNGENVLYS SANIPNRHRQ SNELLEELSA AGGASAEPNV FGKH

P11496; Corazonin. Periplaneta Americana.
QTFQYSRGWT N

(10) Q8IA34; Neuropeptide IFamide preproprotein precursor(IFAMIDE).
MALRFTLTLL LVTILVAAIL LGSSEA**AYRK PPFNGSIF**GK RNSLGKSKIR IPLKPPPISP SRLRQRQNER RLRGGHGGVS HVVSPERQQI GPRPATPPPR TDLEPTTNTP ATGGQMLCLL VRLNVEMPDV KKVMYKIYNV SRAYRYIELM PYIYIQYSIN LQH

P83322; FMRFamide-like neuropeptide FLP7. Penaeus monodon.
GYRKPPFNGS IF

(11) Q8SZ21; Tachykinin (TK) OR CG14734.
MRPLSGLIAL ALLLLLLLTA PSSAADTETE SSGSPLTPGA EEPRRVVKRA PTSSFIGMRG KKDEEHDTSE GNWLGSGPDP LDYADEEADS SYAENGRRLK KAPLAFVGLR GKKFIPINNR LSDVLQSLEE ERLRDSLLQD FFDRVAGRDG SAVGKRAPTG FTGMRGKRPA LLAGDDDAEA DEATELQQKR APVNSFVGMR GKKDVSHQHY KRAALSDSYD LRGKQQRFAD FNSKFVAVRG KKSDLEGNGV GIGEDHEQAL VHPWLYLWGE KR**APNGFLGM R**GKRPALFE

P81733; Tachykinin-related peptide 2. Leucophaea maderae.
APEESPKR**AP SGFLGVR**

(12) Q95NV8; Allatostatin preprohormone(AST2) OR BCDNA:RH36507 OR CG14919
MMKFVQILLC YGLLLTLFFA LSEARPSGAE TGPDSDGLDG
QDAEDVRGAY GGGYDMPAQA IYPNIPMDRL QMLFAQYRPT
YSAYLRSPTY GNVNELYRLP ESKR**QVRYRQ CYFNPISCF**R
K

P42559; Allatostatin. Manduca Sexta.
QVRFRQCYFN PISCF

(13) Q9NIP6; Cardio acceleratory peptide 2b precursor(CAPA OR MT-CAP2B) OR CG15520
MKSMLVHIVL VIFIIAEFST AETDHDKNRR GANMGLYAFP
RVGRSDPSLA NSLRDGLEAG VLDGIYGDAS QEDYNEADFQ
KKASGLVAFP RVGRGDAELR KWAHLLALQQ VLDKR**TGPSA**
SSGLWFGPRL GKRSVDAKSF ADISKGQKEL N

O45027; PBAN-type neuropeptides precursor (PBAN). Mamestra brassicae.
GLWFGPRIGK RSLRMATEDN RQAFFKLLEA ADALKYYYDQ
LPYEMQADEP ETRVTKKVIF TPKLGRSLAY DDKVFENVEF
TPRLGRRLAD DMPATPADQE MYRPDPEQID SRTKYFSPRL
GRTMNFSPRL GRELAYEMVP SKIRVVRSTN KTQST

(14) Q9U4J0; Ecdysis-TRIGGERING hormone(ETH) OR CG18105
MRIITVLSVS LLVGLVAISQ ADDSSPGFFL KITKNVPRLG
KRGENFAIKN LKTIPRIGRS EHSSVTPLLA WLWDLDTSPS
KRRLPAGESP AKEQELNVVQ PVNSNTLLEL LDNNAIPSEQ
VKFVHWKDFD RALQADADLY SKVIQLGRR**P DQHLKQTLSF**
GSFVPIFGDE QNPDFMMYKN NEDQELYGGG NRYDRQFLKY
NIL

Q95336; Progonadoliberin II precursor (GNRH2). Tupaia glis.
MASSMLGFLL LLLLLMAAHP GPSEAQHWSH GWYPGGKRAS
NSPQDPQSAL RPPAPSAAQT AHSFRSAALA SPEDSVPWEG
RTTAGWSLRR **KQHLMRTLLS AAGAPRPAAV PIKP**

(15) Q9VC44; Allatostatin (AST) OR BCDNA:RE16553 OR CG13633
MNSLHAHLLL LAVCCVGYIA SSPVIGQDQR SGDSDADVLL
AADEMADNGG DNIDKRVERY AFGLGRRAYM YTNGGPGMKR
LPVYNFGLGK **RSRPYSFGL**G KRSDYDYDQD NEIDYRVPPA
NYLAAERAVR PGRQNKRTTR PQPFNFGLGR R

P81817; Carcinustatin 14. Carcinus maenas.
YSFGL

(16) Q9VCW0; FLP-1 OR F23B2.5 (CCAP) OR CG4910
MRISLRLLAL LACAICSQAS LERENNEGTN MANHKLSGVI
QWKYEKR**PFC NAFTGC**GRKR TYPSYPPFSL FKRNEVEEKP
YNNEYLSEGL SDLIDINAEP AVENVQKQIM SQAKIFEAIK
EASKEIFRQK NKQKMLQNEK EMQQLEERES K

P38556; Cardioactive peptide. Spodoptera eridania.
PFCNAFTGC

(17) Q9VG55; Hugin protein precursor(HUG) OR CG6371
MCGPSYCTLL LIAASCYILV CSHAKSLQGT SKLDLGNHIS
AGSARGSLSP ASPALSEARQ KRAMGDYKEL TDIIDELEEN
SLAQKASATM QVAAMPPQGQ EFDLDTMPPL TYYLLLQKLR
QLQSNGEPAY RVRTPRLGRS IDSWRLLDAE GATGMAGGEE
AIGGQFMQR**M VKKSVPFKPR L**GKRAQVCGG D

P82618; Pyrokinin-3. Periplaneta Americana.
LVPFRPRL

(18) Q9VIQ0; Short neuropeptide F (SNPF) OR CG13968
MFHLKRELSQ GCALALICLV SLQMQQPAQA EVSSAQGEHL
VQPPPEKQSS KDSFLGTPLS NLYDNLLQRE YAGPVVFPNH
QVERK**AQRSP SLRLRF**GRSD PDMLNSIVEK RWFGDVNQKP
IR**SPSLRLRF** GRRDPSLPQM RRTAYDDLLE RELTLNSQQQ
QQQLGTEPDS DLGADYDGLY ERVVRKPQRL RWGRSVPQFE
ANNADNEQIE RSQWYNSLLN SDKMRRMLVA LQQQYEIPEN
VASYANDEDT DTDLNNDTSE FQREVRKPMR LRWGRSTGKA
PSEQKHTPEE TSSIPPKTQN

P83277; FMRFamide-like neuropeptide FLP4. Macrobrachium rosenbergii.
APALRLRF

(19) Q95SV5; CG13586 OR BCDNA:SD05282.
MCSRNIKISV VLFLVLIPIF AALPHNHNLS KR**SNFFDLEC**
KGIFNKTMFF RLDRICEDCY QLFRETSIHR LCKKDCFDSK
WFGECLKVLL IPEEEISNLQ HFLRVVNGSP ISFNMGPQT

P30814; Crustacean hyperglycemic hormone. Armadillidium vulgare.
RIFDTSCKGF YDRGLFAQLD RVCEDCYNLY RKPHVAAECR
RDCYTTEVFE SCLKDLMMHD FINEYKEMAL MVS

(20) Q9VLK4; Diuretic hormone class-II precursor(DH31) OR CG13094.
MTNRCACFAL AFLLFCLLAI SSIEAAPMPS QSNGGYGGAG
YNELEEVPDD LLMELMTRFG RTIIRARNDL ENSKR**TVDFG**
LARGYSGTQE AKHRMGLAAA NFAGGPGRRR RSETDV

P82372; Diuretic hormone class II. Diploptera punctata.
GLDLGLSRGF SGSQAAKHLM GLAAANYAGG P

(21) Q9VQ66; Neuropeptide like 4 precursor OR CG15361
MFKLLVVVFA ALFAAALA**VP APVARANPAP IPIASPEPAP**
QYYYGASPYA YSGGYYDSPY SYYG

P01193; Corticotropin-lipotropin precursor (POMC OR POM). Mus musculus.
MPRFCYSRSG ALLLALLLQT SIDVWSWCLE SSQCQDLTTE
SNLLACIRAC KLDLSLETPV FPGNGDEQPL TENPRKYVMG
HFRWDRFGPR NSSSAGSAAQ RR**AEEEAVWG DGSPEPSPRE**
GKRSYSMEHF RWGKPVGKKR RPVKVYPNVA ENESAEAFPL
EFKRELEGER PLGLEQVLES DAEKDDGPYR VEHFRWSNPP
KDKRYGGFMT SEKSQTPLVT LFKNAIIKNA HKKGQ

(22) Q9VV28; Neuropeptide like 3 precursor CG13061/CG18502
MFKLCVFVAL LSLAAA**APAP APAPAPAPGL IGPGIVAPGI**
WGPTVVGSPL LAPQVVSVVP GAISHAAITQ VHPSPLLIKS
VHGLGPVVIG

P45646; Lutropin beta chain precursor(LHB). Meleagris gallopavo.
MGGAQVLVLM TLLGTPPVTT GTPPVVVDPS VAVVGPPLGL
GGGGRPPCRP INVTVAVEKD ECPQCMAVTT TACGGYCRTR
EPVYRSPLGR **PPQSSCTYGA LRYERWALWG CPIGSDPRVL**
LPVALSCRCA RCPIATSDCT VQGLGP AFCG APGGFGIGE

(23) Q9VVF7; Allatostatin/MIP OR CG6456
MAHTKTRRTY GFLMVLLILG SACGNLVASG SAGSPPSNEP
GGGGLSEQVV LDQLSESDLY GNNKR**AWQSL QSSW**GKRSSS
GDVSDPDIYM TGHFVPLVIT DGTNTIDWDT FERLASGQSA
QQQQQQPLQQ QSQSGEDFDD LAGEPDVEKR AWKSMNVAWG
KRRQAQGWNK FRGAWGKREP TWNNLKGMWG KRDQWQKLHG
GWGKRSQLPS N

Q95YF7; Prothoracicostatic peptide(PTSP). Bombyx mori.
MRWCLFALWV FGVATVVTAA EEPHHDAAPQ TDNEVDLTED
DKR**AWSSLHS GWA**KRAWQDM SSAWGKRAWQ DLNSAWGKRG
WQDLNSAWGK RAWQDLNSAW GKRGWQDLNS AWGKRDDDEA
MEKKSWQDLN SVWGKRAWQD LNSAWGKRAW QDLNSAWGKR
GWNDISSVWG KRAWQDLNSA WGKRAWQDMS SAWGKRAPEK
WAAFHGSWGK RSSIEPDYEE IDAVEQLVPY QQAPNEEHID
APEKKAWSAL HGTWGKRPVK PMFNNEHSAT TNEA

(24) Q9W0W6; Neuropeptide like 1 precursor (NPLP1)
MQAVLQSAHS SRRLMLLLSM LLNAAIQPRS IIVSATDDVA
NVSPCEMESL INQLMSPSPE YQLHASALR**N QLKNLLRERQ LAVGEEQPLG EYPDYLEED**K RSVAALAAQG LLNAPKRSLA
TLAKNGQLPT AEPGEDYGDA DSGEPSEQKR YIGSLARAGG
LMTYGKRNVG TLARDFQLPI PNGKRNIATM ARLQSAPSTH
RDPKRNVAAV ARYNSQHGHI QRAGAEKRNL GALKSSPVHG
VQQKREDEEM LLPAAAPDYA DPMQSYWWYP SYAGYADLDW
NDYRRAEKRF LGRVLPPTRA TASTHRSRL

P01283; Vasoactive intestinal peptide(VIP). Rattus norvegicus.
MESRSKPQFL AILTLFSVLF SQSLAWPLYG PPSSVRLDDR
LQFEGAGDPD QVSLKADSDI LQNALAENDT PYYDVSRNAR
HADGVFTSDY SRLLGQISAK KYLESLIGKR ISSSISEDPV
PVKRHSDAVF TDNYTRLRKQ MAVKKYLNSI LNGKR**SSEGD SPDFLEELEK**

(25) Q9Y1K3; Neuropeptide F (NPF) OR CG10342
MCQTMRCILV ACVALALLAA GCRVEASNSR PPRKNDVNTM
ADAYKFLQDL DTYYGDRARV RFGKR**GSLME ILRNHEMDNI NL**GKNANNGG EFVSIE

P23442; Islet amyloid polypeptide precursor(IAPP).Mesocricetus auratus.
MHISKLPAAL LIFSVALNHL KATPVRSGTN HQMDKRKCNT
ATCATQRLAN FLVHSNNNLG PVLSPTNVGS NTYGKRSAAE
IPDGDSLDLF LL

(26) Q9VT50; Probable insulin-like peptide 1 precursor (ILP1) OR CG141
MFSQHNGAAV HGLRLQSLLI AAMLTAAMAM VTPTGSGHQL
LPPGNHKLCG PALSDAMDVV CPHGFNTLPR KRESLLGNSD
DDEDTEQEVQ DDSSMWQTLD GAGYSFSPLL TNLYGSEVLI
KMRRHRR**HLT GGVYDECCVK TCSYLELAIY CLPK**

(27) Q9VT51; Probable insulin-like peptide 2 precursor(IRP) OR CG8167
MSKPLSFISM VAVILLASST VKLAQGTLCS EKLNEVLSMV
CEEYNPVIPH KRAMPGADSD LDALNPLQFV QEFEEEDNSI
SEPLRSALFP GSYLGGVLNS LAEVRR**RTRQ RQGIVERCC**K
KSCDMKALRE YCSVVRN

P01325; Insulin 1 precursor(INS1 OR INS-1). Mus musculus.
MALLVHFLPL LALLALWEPK PTQAFVKQHL CGPHLVEALY
LVCGERGFFY TPKSRREVED PQVEQLELGG SPGDLQTLAL
EVARQKR**GIV DQCCTSICSL YQLENYCN**

(28) Q9VT52; Probable insulin-like peptide 3 precursor(ILP3) OR CG14167).
MGIEMRCQDR RILLPSLLLL ILMIGGVQAT MKLCGRKLPE
TLSKLCVYGF NAMTKRTLDP VNFNQIDGFE DRSLLERLLS
DSSVQMLKTR R**LRDGVFDEC CLKSCTMDEV LRYCAAKPRT VTCNKL**

P26733; Bombyxin B-1 precursor(BBXB1). Bombyx mori.
MKTSVMFMLV IVISLMCSGE AQEVARTYCG RHLADTLADL
CFGVEKRGGA QYAPYFWTRQ YLGSRGKR**GV VDECCFRPCT LDVLLSYCG**

(29) Q9VT53; Probable insulin-like peptide 4 (ILP4) OR CG6736
MSLIRLGLAL LLLLATVSQL LQPVQGRRKM CGEALIQALD
VICVNGFTRR VRRSSASKDA RVRDLIRKLQ QPDEDIEQET
ETGRLKQKHT DADTEKGVPP AVGSGRKLRR HRR**RIAHECC KEGCTYDDIL DYCA**

Q9TRM8; Prorelaxin precursor(RLN). Canis familiaris.
MLRWFLSHLL GVWLLLSQLP REIPATDDKK LKACGRDYVR
LQIEVCGSIW WGRKAGQLRE RRQISEPLAE VVPSSIINDP
EILSLMLQSI PGMPQELRIA TRSGKEKLLR ELHFVLEDSN
LNLEEMKKTF LNTQFEAEDK SLSKLDKHPR KK**RDNYIKMS DKCCNVGCT**R RELASRC

(30) Q9W4Q9; CG13317 protein (ILP7) OR CG13317
MTRMIIQNSG SWTLCGAVLL FVLPLIPTPE ALQHTEEGLE
MLFRERSQSD WENVWHQETH SRCRDKLVRQ LYWACEKDIY
RLTRRNKKRT GNDEAWIKKT TTEPDGSTWL HVNYANMFLR
SRRSDGNTPS ISNECCTKAG CTWEEYAEYC PSNKRRNHY

AAQ89696; Insulin-like peptide 5 precursor. Anopheles gambiae.
MWLPLALCVL LEFADIVSAS GGLDDALEVT FSERTRADWE
KVWHQESHSR CREKLIRHLY WACEKDIYRI SRRSGDGNGI
AGMMEKRTSM VDEGQLVPYP WAIDREVAYA FLRTRR**TGKR RSGGSITAEC CTRTGCTWEE YAEYCPSN**KR LNQYRRRK

(31) P05623; Accessory gland-specific peptide 70A precursor(ACP70A OR PAPB) OR CG17673.
MKTLALFLVL VCVLGLVQSW EWPWNRK**PTK FPIPSPNPRD KWCRLNLGPA WGGRC**

O18417; Accessory gland-specific peptide 70A precursor(PAPB OR ACP70A). Drosophila sechellia.
MKTLSVFLVL VCLLGLVQSW EWPWNRQPTR YPIPSPNPRD KWCRLNLGPA WGGRC

(32) P10333; Accessory gland-specific peptide 26Aa precursor(ACP26AA OR MST26AA OR MST355A) OR CG8982.
MNQILLCSPI LLLLFTVASC DSEQQLDSAM HLKSDSTKSA
SLKNVAPKND ETQAKIAKDD VALKDAKKGD YIMDIDISDL
PLDDYPINRS KSLKSSSIDL NNIPFNKGLD DFPAKEKNQG
SNQSALKALQ QRLLTEQNNS LLLR**NHSIYL MKEIEA**RKTD
IIKVRQLNLD LELELNTVNR RLLELNGQLQ NTRKSTKPCK
KRSSKDSAPP AANQFQEANV RNTYRNKYLT LLKELSQKIN
NEIAKVATDV PTETNPSQGN LPTL

P33735; Accessory gland-specific peptide 26Aa precursor(ACP26AA OR MST26AA OR MST355A). Drosophila mauritiana.
MNQILLCSQI LLLFFTVANC DGEHQLDSSV DLKSAVLKNV
APKNVATQAE IVKDDVALKS GKKGDYVMDI EVSDMPLDDY
PINNSKSRKN SSTLPSPILT DKLNQGSNQI ALKALKHRLV
MEQNNNLFLR **NHSVSLMNEI EA**RKTDIIQA RQLNIDLELE
LESLKRKLSE MNVQNARKST KSCKKRPSKD IAPPVNQLQE
VIVKNTYRNK YLTLLTQLAQ KINYEIANVN NPATDVPTGK
SPSEGNPSTT

(33) P10334; Accessory gland-specific peptide 26Ab precursor(ACP26AB OR MST26AB OR MST355B OR CG9024
MNYFAVICIF SCICLWQFSD A**APFISVQSS SQS**RSQKVMN GMLRTLYDYS VQDSVNDATG HLIQTHKADF NSDVMSPDEI ESVRQQLNMA

P33738; Accessory gland-specific peptide 26Ab precursor(ACP26AB OR MST26AB OR MST355B). Drosophila mauritiana.
MNYFAVLCIF SCICLWQFSD AAPFISVQSS SQSRSQKVMN GMLRTLYDYS VQDSVNDATG HLIHTHKADF NSDVMSPEEI ERVRQQLNMA

(34) P16548; Accessory gland-specific peptide (ACP95EF OR MST95E OR MSP316) OR CG17924
MASVKLFFIA ILVVALSLNT SAAVLNPSST AKPRFETKDR K**LSAGALQSL A**G

Q62542; Insulin 1(INS1). Mus spretus.
GSPGDLQTLA LEVARQKRGI VDQCCTSICS LYQLENYCN

(35) P24492; Diptericin precursor(DPT) OR DIPT OR CG12763
MQFTIAVALL CCAIASTLAY PMPDDMTMKP TPP PQYPLNL QGGGGGQSGD GFGFAVQGHQ KVWTSDNGR**H EIGLNGGYGQ HLGGPYGNSE PSWKVGSTYT YRFPNF**

Q8WTD5; Antimicrobial peptide diptericin DipA.Glossina morsitans.
PQSPPAQIKD PKIYASGGGS PKDGYNVNVD VRKNVWVSQN GR**HSIDATGG YSQHLGGPYG NSRPDFRGGA SYTYRF**

(36) Q9V8P6; Dipterecin B precursor (BCDNA:RH29451) or CG10794
MHFTASLLFI GLACAFSSAW AYPYPDPREI VNLQPEPLAY APNFDVPLHR VRRQFQLNGG GGGSPKQGFD LSLNGRAPVW QSPNGR**HSFD ATGSYAQHLG GPYGNSRPQW GAGGVYTFRF**

Q8WTD5; Antimicrobial peptide diptericin DipA.Glossina morsitans.
PQSPPAQIKD PKIYASGGGS PKDGYNVNVD VRKNVWVSQN GR**HSIDATGG YSQHLGGPYG NSRPDFRGGA SYTYRF**

(37) P36193; Drosocin precursor (DRO) OR CG10816
MKFTIVFLLL ACVFAMAVAT PGKPRPYSPR PTSHPRPIRV RR**EALAIEDH LAQAAIRPPP ILPA**

P10419; Antho-RFamide neuropeptides type 1 precursor. Anthopleura elegantissima.
MTTVSYVTIL LTVLVQVLTS DAKATNNKRE LSSGLKERSL SDDAPQFWKG RFSRSEEDPQ FWKGRFSDPQ FWKGRFSDPQ FWKGRFSDPQ FWKGRFSDPQ FWKGRFSDPQ FWKGRFSDPQ FWKGRFSDGT KRENDPQYWK GRFSRSFEDQ PDSEAQFWKG RFARTSSGEK REPQYWKGRF SRDSVPGRYG RELQGRFGRE LQGRFGREAQ GRFGRELQGR FGREFQGRFG REDQGRFGRE DQGRFGREDQ GRFGREDQGR FGREDQGRFG REDQGRFGRE LQGRFGREFQ GRFGREDQGR FGREDQGRFG RELQGRFGRE DQGRFGREDQ GRFGREDLAK EDQGRFGR**ED LAKEDQ**GRFG REDIAEADQG RFGRNAAAAA AAAAAAKKRT IDVIDIESDP KPQTRFRDGK DMQEKRKVEK KDKIEKSDDA LAKTS

(38) P14954; Cecropin A1/A2 precursor (CECA1) OR CG1365 AND CECA2 OR CG1367. Drosophila melanogaster.
MNFYNIFVFV ALILAITIGQ SEA**GWLKKIG K**KIERVGQHT RDATIQGLGI AQQAANVAAT ARG

P49931; Antibacterial peptide PMAP-36 precursor(PMAP36).Sus scrofa.
METQRASLCL GRWSLWLLLL GLVVPSASAQ ALSYREAVLR AVDRLNEQSS EANLYRLLEL DQPPKADEDP GTPKPVSFTV KETVCPRPTW RPPELCDFKE NGRVKQCVGT VTLNPSNDPL DINCDEIQSV GRFRRLRKKT RKR**LKKIGKV LKWIPPIVGS IPLGCG**

(39) P45884; Attacin A precursor(ATTA) OR CG10146
MQKTSILIVA LVALFAITEA LPSLPTTGPI RVRRQVLGGS LTSNPAGGAD ARLDLTKGIG NPNHNVVGQV FAAGNTQSGP VTTGGTLAYN NAGHGASLTK THTPGVKDVF QQEAHANLFN NGRHNLDAKV FASQNKLANG FEFQRNGAGL DYSHINGHGA SLTHSNFPGI GQQLGLDGR**A NLWSSPNRAT TLDLTGSASK WTSGPFANQK PNFGAGLGLS HHF**G

(40) Q9V751;Attacin B precursor(ATTB) OR ATTB1 OR CG18372
MQKTSILILA LFAIAEAVPT TGPIRVRRQV LGGSLASNPA GGADARLNLS KGIGNPNHNV VGQVFAAGNT QSGPVTTGGT LAYNNAGHGA SLTKTHTPGV KDVFQQEAHA NLFNNGRHNL DAKVFASQNK LANGFEFQRN GAGLDYSHIN GHGASLTHSN FPGIGQQLGL DGR**ANLWSSP NRATTLDLTG SASKWTSGPF ANQKPNFGAG LGLSHHF**G

Q8WTD3; Antimicrobial peptide attacin AttA. Glossina morsitans.
MQSFKICFFI SCLSVVLVKG QFGGTVSSNP NGGLDVNARL SKTIGDPNAN VVGGVFAAGN TDGGPATRGA FLAANKDGHG LSLQHSKTDN FGSSLTSSAH AHLFNDKTHK LDANAFHSRT HLDNGFKFDR VGGGLRYDHV TGHGASLTAS RIPQLDMNTL GLTGK**ANLWS SPNRATTLDL TGGVSKHFGG PFDGQTNKQI GLGLNSR**F

(41) P82706; Immune-induced protein 1 precursor(IM1) OR CG18108)
MKFFSVVTVF VLGLLAVANA**V PLSPDPGNV IINGDCRVCN VHG**GK

Q9V8G0; Immune-induced peptide 3 precursor(IM3 OR BCDNA:RH58911 OR CG16844). Drosophila melanogaster.
MKFLSLAFVL GLLALANATP LNPGNVIING DCRVCNVRA

(42) O77150; Immune-induced protein 2 precursor(IM2) OR BCDNA:RH08291 OR CG18106.
MKFFSVVTVF VFGLLALANA **VPLSPDPGNV VINGDCKYCN VHG**GK

Q9V8G0; Immune-induced peptide 3 precursor(IM3) OR BCDNA:RH58911 OR CG16844.
MKFLSLAFVL GLLALANATP LNPGNVIING DCRVCNVRA

(43) Q9VLV9;PROCT OR CG7105.CG7105 protein
MGVPRSHGTG IGCGSGHRWL LVWMTVLLLV VPPHLVDGR**Y LPTRSHGDDL DKLRELMLQI LELSNEDPQQ QQQQQQQQQH PQLRLHNEAT GGSSSSSNIN NPRVSNGNSN AAWLQKLSAM GALDELGGDG ARFGPNYGRY**

Q8MMJ7;Cytotoxic linear peptide IsCT precursor.Opisthacanthus madagascariensis.
MKTQFAILLV ALVLFQMFAQ SDAILGKIWE GIKSLFGKR**G LSDLDGLDEL FDGEISKADR DFLRELMR**

The putative peptide precursors (28)

(1) P02841; Salivary glue protein Sgs-7 precursor (SGS7) or CG18087
MKLIAVTIIA CILLIGFSDL ALG**GACECQP CGPG**GKACTG
CPEKPQLCQQ LISDIRNLQQ KIRKCVCGEP QWMI

P35455; Vasopressin-neurophysin 2-copeptin precursor(AVP). Mus musculus.
MLARMLNTTL SACFLSLLAF SSACYFQNCP RGGKR**AISDM**
ELRQCLPCGP GGKGRCFGPS ICCADELGCF VGTAEALRCQ
EENYLPSPCQ SGQKPCGSGG RCAAVGICCS DESCVAEPEC
HDGFFRLTRA REPSNATQLD GPARALLLRL VQLAGTRESV
DSAKPRVY

(2) P07701; Salivary glue protein Sgs-5 precursor(SGS5) OR CG7596.
MFNIKLLLLL LAVSWFHHGQ AVQETKIEEK PVSEPEIESE
IKNSTSVPSK CNIYYRNYQW ALQDCVCRCF QNECLMQIES
DQRK**KEGRSP FVPVTEELCR SFICKKCSVG FPVVAEFPIP**
APCGCNRKPG SIATERFYSL CHLLKFSAEN SKPFLTYSYC
WPF

Q7ZZZ3; Putative growth hormone like protein-1(YGHL1). Xenopus laevis.
MAQPGDVLPT YDMSDSQTSK LIRK**SKESPF VPIGMAGFAA**
VVAFGLFKLK SRGNTKMSVH LIHMRVAAQG FVVGAMTCGV
LYSMYKEYLA KPSEH

(3) Q27241; EIG71ED protein (EIG71ED) OR L71-4 OR CG7350
MHTTAVVTLF SVLLVVLVAG QNRNCDELTR RCERCVETLN
NAADRNLPVL NQECRTKTRN NWRWRNVGRC ELTRLNCLGS
NRR**MNCNDIA ELAGMDRIN**

O35417; Beta-neoendorphin-dynorphin precursor(PDYN). Mus musculus.
MAWSRLMLAA CLLVMPSNVM ADCLSLCSLC AVRIQDGPRP
INPLICSLEC QDLVPPSEEW ETCRGFSSFL TLTVSGLRGK
DDLEDEVALE EGISAHAKLL EPVLKELEKS RLLTSVPEEK
FRGLSSSFGN GK**ESELAGAD RMNDEAAQ**GR TVHFNEEDLR
KQAKRYGGFL RKYPKRSSEM ARDEDGGQDG DQVGHEDLYK
RYGGFLRRIR PKLKWDNQKR YGGFLRRQFK
VVTRSQENPN TYSEDLDV

(4) Q00805; Giant-lens protein precursor (ARGOS) OR AOS OR GIL
MPTTLMLLPC MLLLLLTAAA VAVGGTRLPL EVFEITPTTS
TADKHKSLQY TVVYDAKDIS GAAAATGVAS STVKPATEQL
TVVSISSTAA AEKDLAESRR HARQMLQKQQ QHRSIIGGKH
GDRDVRILYQ VGDSEEDLPV CAPNAVCSKI DLYETPWIER
QCRCPESNRM PNNVIIHHHS HSSGSVDSLK YRNYYEREKM
MQHKR**MLLGE FQD**KKFESLH MKKLMQKLGA VYEDDLDHLD
QSPDYNDALP YAEVQDNEFP RGSAHMRHSG HRGSKEPATT
FIGGCPSSLG VEDGHTIADK TRHYKMCQPV HKLPVCKHFR
DYTWTLTTAA ELNVTEQIVH CRCPRNSVTY LTKREPIGND
SPGYRYLFAC SPLTRLRCQR KQPCKLFTVR KRQEFLDEVN
INSLCQCPKG HRCPSHHTQS GVIAGESFLE DNIQTYSGYC
MAND

P01277; Glucagon precursor. Meleagris gallopavo.
MKMKSIYFIA GLLLMIVQGS WQNPLQDTEE KSRSFKASQS
EPLDESRQLN EVKRHSQGTF TSDYSKYLDS RRAQDFVQWL
MSTKRNGQQG QEDKENDKFP DQLSSNAISK RHSEFERHAE
GTYTSDITSY LEGQAAKEFI AWLVNGRGRR DFPEKALMAE
EMGRRHADGT FTSDINKILD DMAAKEFLKW LINTKVTQR**D**
LLGEYQ

(5) Q24155; Trunk protein precursor(TRK) OR CG5619
MKSQSELAIV LTWLAVLGTA QDDADYCAEL STQSLAKILG
QAFNPRYMSI DPPGEPEEKS YHLGYKRSSY ELPFYADSDA
ISVSHFPAWE TNHFALVEKK KEAPRSKSLR TR**SAFMDRVG**
HPRIDGFKQR PWECSSKINW IDLGLNYFPR YIRSIECIAR
KCWYDHFNCK PKSFTIKVLR RKTGSCIRIN DKLILITAEK
FENDYTQLWI WEEIAVNFCC ECVMLY

Q7Z1Z6; Prothoracicotropic hormone. Helicoverpa armigera.
MITRPLLCVI VCFGFFILIQ SLVPKVMAMK HSNVDEYMLE
DQRTRKRKNY VVRLARDSEI LGKPGNVGTN YDTDSFQLEP
ANPDELSAFI VDYANMIRND VILLDKSVET RTRKR**GNIKV**
EKYNNQALPD PPCACKFSPN RTDLGENTYP RYIETRNCSQ
ARQQSCRPPY VCRENYYNIT IIRRKEFQNQ ATLEDIPHDL
KFRWVAENYP VSVGCVCTRD YYATEK

(6) Q8IME0; CG32851 CG15068
MKLLSITFLF GLLALASA**NP LSPGNVIING DCKVCNIRGD**

(7) Q9V8F7; CG18107
MRFFAIVTVF VLGLLALANA**I PLSPDPGNV IINGDCVNCN**
VRGGK

(8) Q9V8G2; CG15065
MKWMSLVFLC GLLAMAVA**SP LNPGNVIING DCRHCNVRGG**

(9) Q9VD48; CG5791 OR BCDNA:RH16331
MKYLTCVLLP LALIPTLIGA HPSTVVVNGV CLTCPNPNGE
PVYLDGQQYR SFSSSPGDGN VVISRGNDGS GGGGGTIYRR
GGNTIVNGRC QHCNVDPY

Q9V8G0; Immune-induced peptide 3 precursor(IM3 OR BCDNA:RH58911 OR CG16844). Drosophila melanogaster.
MKFLSLAFVL GLLALANATP LNPGNVIING DCRVCNVRA

(10) Q8IMR6; CG31081
MLGIVFLTLL AGSSA**ELGYQ YQQNSYGGPV NSYGNEAVLG**
DERYHSQPGN HYQENADFHK HFYAFEAPYD SVEEVDLAET
KLSSLAQKNL QVVFIKAPEN KAVVGALNAL AKQTSEDKTA
IYVLNKQTDV NELASQLSAL KAHHKHKPQV HFVKYKTEEE
AAQAQQYIQA QYGGGSSIPQ PGKASSLGYY PEQQPQYEQD
APSEEYPAGQ VGYLPSPQQS AYQPQSGYLP PLPSYSSISQ
GYNAAGSSAG VSTIGQIDLP PVPEAQQDLT GNYNNAGVDY
RSARSRRVDF RANERHRRGS RMVFPSANPG KRLRL

Accession:Q16992; LWamide neuropeptides precursor. Anthopleura elegantissima.
MALKCHLVLL AITLLLAQCS GSVDKKDSTT NHLDEKKTDS
TEAHIVQETD ALKENSYLGA EEESKEEDKK RSAAPQQPGL
WGKRQKIGLW GRSADAGQPG LWGKRQSPGL WGRSADAGQP
GLWGKRQNPG LWGRSADAGQ PGLWGKRQNP GLWGRSADAG
QPGLWGKRQN PGLWGRSADA RQPGLWGKRE IYALWGGKRQ
NPGLWGRSAD PGQPGLWGKR ELVGLWGGKR QNPGLWGRSA
EAGQPGLWGK RQKIGLWGRS ADPLQPGLWG KRQNPGLWGR
SADPQQPGLW GKRQNPGLWG RSADPQQPGL WGKRQNPGLW
GRSADPQQPG LWGKRQNPGL WGRSADPQQP GLWGKSPGLW
GRSADPQQPG LWGKRQNPGF WGRSADPQQP GLWGKRQNPG
LWGRSADPQQ PGLWGKRQNP GLWGRSADPQ QPGLWGKRQN
PGLWGRSADP QQPGLWGKRQ NPGLWGRSAG SGQLGLWGKR
QSRIGLWGRS AEPPQFEDLE DLKK**KSAIPQ PKGQ**

(11) Q8MS86; BCDNA:LP04693
MSCSAWTQTP THTHKHRAIQ IVTIISVLII ECSALVACSL
TPTSSLPALH RR**WKILTAGS HFRWL**

P11885;Corticotropin-lipotropin precursor.
Rana catesbeiana.
MLQPVWHACI LAILGVFIFH VGEVRSQCWE SNKCTDLSSE
DGILECIKAC KMDLSAESPV FPGNGHIQPL SENIRK**YVMS**
HFRWNKFGRR NSTSNDNNNN NGGYKREDIA NYPILNLFLG
SDNQNTQEGI MEDDALDRQD SKRSYSMEHF RWGKPVGKKR
RPIKVFPTDA EEESSESFPI ELRRELSLEF DYPDTNSEEE
LDNGELLEGP VKKGRKYKMH HFRWEGPPKD KRYGGFMTPE
RSQTPLMTLF KNAIIKNAHK KGQ

(12) Q8MVX6; Odorant binding protein b(OBP-B)
MRVLLAFVLL LGLSVLATK**E PEEVKIVSEC AKENNVH**RKK
ALDLLMSYRL KKKTHNVMCF INCIFERTNI LQKVKEKVVK
ENHNCDSIKD ADKCAESFQK FQCLVKIEMK VRGIDRG

P01272; Glucagon precursor(GCG).Bos Taurus.
MKSLYFVAGL FVMLVQGSWQ RSLQNTEEKS SSFPAPQTDP
LGDPDQINED KRHSQGTFTS DYSKYLDSRR AQDFVQWLMN
TKRNKNNIAK RHDEFERHAE GTFTSDVSSY LEGQAAKEFI
AWLVKGRGRR **DFPEEVNIVE EL**RRRHADGS FSDEMNTVLD
SLATRDFINW LLQTKITDRK

(13) Q95U18; GH13848p(CG14995).
MGLRSGVWGL GDLPTLLLTS VAPAAFTRPN SRGNLLTSLE
VSTLERVTIS WTGKTRAPRE KRGSDLSDVS IIKRMRGVEV
LALSVNKIST LSTFEDCTKL QELYLRKNSI SDINEIAYLQ
NLPSLRNLWL EENPCCERAG PNYRSIVLRA LPNLKKLDNV
EVTQQEVEDA LRGGGVAAPE DEVYEDAYQQ QQQSRRSSPQ
QILQQQQHSY PQHSPPPQQQ YQQQQQQQQQ QQR**GCTTPTK**
EEYYQSDRPA YPAHYRHSQT DLTEWEEHQQ VPQVHHNPYG
SQKQLHQPQR RSAGPEMTPY RNGSARENGG EWDPEDRSRA
RRPEGRYSDG TSSLSASVMN HYSGYHRRPI NRNSNILSAA
LCLVKELDYA SLEVLEHAVR CRIDELANE

Q8JGU0; Neuropeptide FF-related PQRF(PQRF).
Brachydanio rerio.
MNGLLEDRLL VEMLR**SLLHG SQRYERNPSV LHQPQRF**GRG
ARSGLSTEER IQSRDWETVP GQIWSMAVPQ RFGKK

(14) Q9V7Q4; CG15712.
MTFSTSVSVI LLISAISCWA TAFGGLREDL KDFVALVPRR
RIGFIAARYY IFDPKFRQAV EFVRSDEFIA TWQQVRATPD
FVNIINYVSD YGSGYDITTL VDSLPTRLRA YQLSRTVPVE
LMLRRDLNTF LWDVIHSLPR TRIYSLIAQK SKQSTEFAK**L**
YKALRDKEFK ELVQRARLSR DLQAPIKKLS QKSINVDEIL
QIVFEVISWG PKTS

BAB23997; Adult female placenta cDNA, RIKEN
full-length enriched library,clone:
1600002A15 product:growth.Mus musculus
MLLWVLFVIL ILTSGSHCSL PPSPPFRMQR HVDAIFTTNY
RKLLSQLYAR K**VIQDIMNKQ GERIQEQRAR LS**RQEDSMWT
EDKQMTLESI LQGFPRMKPS ADA

(15) Q9VFI5; CG8087.
MKATTILAVV SVLTACLLRS SEAVTCTADA TVTGCIDCTT
NPTDSECVAE AAADTTSTTV ATPTTTATTT SATATTTAAS
STNTSSGRRK IVRITNLRYT NVRRIRVNRN GSGSTTVRNR
RRRNNSRRVN VRR**ANGNVIV VG**

Q25160; Putative neuropeptide(HDS2).
Helix lucorum.
MGRVRLCLCV SLVISCLAQV GMFVPIILDV DLDFDTNPLL
KVADGVLEPM DTNDLTVLPS RSRRTLPPWL SGTGQSNRNL
TNVHPLMNRR **FRNNNVIVVA HS**

(16) Q9VIQ6; CG16772
MFIKLLLISQ LLALSYAQLS LDEAKKQIDA SLYSEEETTD
DTHLPEPHLP PQYQPHHPHK KVTTTTPEPE TTTPKPETTT
KPAIEQEGEA TAAPDDGLGQ LDDGLQQNND GSALPESTTP
EPETTTPTTT TTTTPKPHKH EPHFHPHSYP HPHPYPYPIQ
YPYPHPGVIF SAAGPKLTPP STTPPTAKDA DKESPELSGY
PSYPTFRSPY SPYQPQVFPQ PRNWPSFPGY GPPSPGFGYP
HNHNHDHEED SGEDKPKDKS GEKDKDEDVA LKPPGFSYPQ
VYLIPRR**PVI SVPSFPRPGG GYGSPYGYGL Y**

P81819; Carcinustatin 16. Carcinus maenas.
GGPYSYGL

(17) Q9W1F8; CG13565
MNLYVLLAVV SVFLNFIHAA PGVDISNDEL LDGKYLCEAG
SKKYDGPFIV RLISAANGQT VVCYECSQSE FKTKYSVKQC
AAGKIGSGHH RDLVPYLVRM DPLYKDTWSS KLKR**NFDEID**
KASASFSILN QLV

P37086; Orcokinin. Orconectes limosus.
NFDEIDRSGF GFN

(18) Q9VVW7; CG11577
MLLKRLPGFV TLWVVLQLAG ADSPEEEQGV RYANRCEACK
ILATELEARL GETGKSHDVI EIGYSVDDVK PKKRTEYRRS
ELRLLESLEN VCERVLEYNL HKERSDSTRF AKGMSQTFQT
LHGLVDKGVK VDLGIPYELW DKPPVEVTQM KTQCENLLEE
YEETISEWYF KHQDEKSLKK HLCEDHVLKK **KAERECLKEQ**
LAPPEAKKAK REKAKGDKEE L

Q7TSQ5; Gastrin-releasing peptide.
Arvicanthis ansorgei.
QAPRGPAAPV STGAGGGTVL AKMYPRGSHW AVGHLMGKK**S**
TDESLYAADR DGLKEQLRGH VPWEEAARNL LGLLEATGNR
SHQPPQHQPL GSLPPTWDPE DGSYFHDVQN AKLVDSVLQV
LKGKEGTAS

(19) BK002023; HDC09365.
MMFRSVIPVL LFLIPLLLSA QAANSLRACG PALMDMLRVA
CPNGFNSMFA KRGTLGLFDY EDHLADLDSS ESHHMNSLSS
IRR**DFRGVVD SCC**RKSCSFS TLRAYCDS

P15411; Bombyxin A-2 precursor(BBXA2).
Bombyx mori.
MKILLAIALM LSTVMWVSTQ QPQEVHTYCG RHLARTMADL
CWEEGVDKRS DAQFASYGSA WLMPYSAGR**G IVDECCLRPC**
SVDVLLSYC

(20) BK002187; HDC10589.
MQFIYLTLAL GLIFTTALQA AIIPLTLIKN GIEANSQALP
TDTEKFGYLE FKPNGSLILR R**APNQSGSNL QDLVML**RGVL
QALKASPSKM SDIGGETRLS LRIYGDGVEH KFPPILENII
QRIQTYFSVY RFTDTSKPGG LQRIELTTQP PDDSPDVTTA
KAENDDELIA VGEQDAYITV GDDTD

Q805D8; Atrial natriuretic peptide(ANP).
Fugu rubripes.
MTALVLWGLL LLLGQHTQVN SHVLGRPFSA SDSSQLKSLL
ERLEETISEA DQEQNPELDQ EVEYDIRDQD PGQRWNLDLG
RDQDQVTATR SEIHSRPSVQ R**SHLQDLLMS L**RKRASSCFG
ARMDRIGNAS GLGCNNGRG

(21) BK002297; HDC11617.
MVSQEFTHLV SIFWLSYLPK SLLSYGHDGH GLIQIDSSFI
NKKQLRKRFT DRKPRQAYSA SQLERLENEF NLDKYLSVSK
RVELSKSLSL TEVQRAVVSL SSFVQREVLV TCSDQPHTTV
SRSLAPYLSG LPLCSFKVRT ITVRPLCRLE SKVHRGPGNL
RQVGDAPGR**L KKIGLCDYWA NKT**KRIMEPW DCG

P82286; Bombinin-like peptides 2 precursor. Bombina variegata.
MNFKYIVAVS ILIASAYARR EENNIQSLSQ RDVLEEESLR
EIRGIGASIL SAGKSALKGF AKGLAEHFAN GKRTAEDHEM
MKRLEAAVRD LDSLEHPEEA SEKETRGFNQ EEKEKR**IIGP**
VLGLVGSALG GLLKKIG

(22) BK002714; HDC14730.
MLLLLLVLLC LLLLLLLLLV LRFCVFAALA TQGPNQTRTV
DELCVQQLLL PGVIRAIAQL AQAARPQAQW AIVVQFTHTR
RMRNVGMAGE PSERPRSRRN LLTQRRSNNL LTSR**DTATNA**
SQPDGRKDGG HFAICYTADR QRRPVQRQEI RPEARSQRDK
QTRGKRSKRD ILLSCSCKFL LSFFGNTKGE SQDEVTFSLI
VELLSLRESL ARLKPFIATM INDQY

P55099; Neurokinin B precursor(TAC3 OR NKNB OR TAC2). Mus musculus.
MRSAMLFAAV LALSLAWTFG AVCEEPQEQG GRLSKDSDLY
QLPPSLLRRL YDSRPVSLEG LLKVLSKASV GPKETSLPQK
RDMHDFFVGL MGKR**NSQPDT PTDVVEENTP SFGILK**

(23) BK002770; HDC15079.
MQAIFILCAL LVCLLVLLRL SWSGGSSESQ DVTTTNPKFL
VHKDSHSSYD WSNDEIRNRA KDNDLNVITG EELLKSKARF
QKLLSILRR**Q PNKDNYSIDL PGQFDDTSDP SNAEDTTIFC**
NPILDENCYP DSIGPRLSAT LLKILA

O35417; Beta-neoendorphin-dynorphin precursor(PDYN).Mus musculus.
MAWSRLMLAA CLLVMPSNVM ADCLSLCSLC AVRIQDGPRP
INPLICSLEC QDLVPPSEEW ETCRGFSSFL TLTVSGLRGK
DDLEDEVALE EGISAHAKLL EPVLKELEKS RLLTSVPEEK
FRGLSSSFGN GKESELAGAD RMNDEAAQGR TVHFNEEDLR
KQAKRYGGFL RKYPKRSSEM ARDEDGGQDG DQVGHEDLYK
RYGGFLRRIR PKLKWDNQKR YGGFLRRQFK VVTR**SQENPN**
TYSEDLDV

(24) BK003312; HDC00783.
MDTSIVIVIV IVIAIA**IDFD IDIPGLQLEL LLSRILGL**KR
QNQGSNKAAT RTATHRTTPK DDKD

Q28318; Prolactin precursor(PRL). Capra hircus.
MDSKGSAQKG SRLLLLLVVS NLLLCQGVVS TPVCPNGPGN
CQVSLRDLFD RAVMVSHYIH NLSSEMFNEF DKR**YAQGKGY**
ITMALNSCHT SSLPTPEDKE QAQQTHHEVL MSLILGLLRS
WNDPLYHLVT EVRGMKGVPD AILSRAIEIE EENKRLLEGM
EMILGQVIPG AKETEPYPVW SGLPSLQTKD EEARHSAFYN
LLHCLRRDSS KIDTYLKLLN CRIIYNNNC

(25) BK003730; HDC05827.
MWPIVMALIR RNAVYITLPI AGVVG**FIGYN IESWISDKYT**
PYSPSIQELR AKRLTEESLN TDAANVEKLR LSSPVLERNL
SPSLQPKA

Q866H2; Parathyroid hormone-like hormone. Sus scrofa.
MLWRLVQQWS VAVFLLSYSV PSCGRSVEEL GRRLKRAVSE
HQLLHDKGK**S IQDLR**RRFFL HHLIAEIHTA EIRATSEVSP
NPKPAPNTKN HPVRFGSDDE GRYLTQETNK VETYKEQPLK
TPGKKKKGKP GKRKEQEKKK RRTRSAWLNC SMVGSGLEVD
HVSDDSETSL ELNSRRH

(26) Q9VNV8; MSOPA OR CG14560
MNFIQIAVLF VLVAVALARP QEDPANLPAP EAAAAPPAAA
AAPPAAAAAP PAPPAPPAAA PQAAPAGGSG RKK**NVNHNVI**
TIG

Q9VBL6; Accessory gland-specific peptide 57Da precursor(MST57DA OR BCDNA:GH19893 OR CG9074). Drosophila melanogaster.
MKFLALFVTL LVVLALVSAQ KSQNTNHNVI VIGAKKPGAA
PAAAAAAAPA APPAAAPAAA PAAPEAGLAD APAES

(27) Q9VDF8; CG31189
MDRLISLTFL CWCIPVMISG ASLRAWVFNV EKCHFGDSTC
LVRSINALIK HYPKGIPEIG LPPLDAYNFP DSVIMESPSR
GPIWMDFRMR DNVNKGFNNA TITHVEGFLY EPNQKQIVLK
VRLPRLVHEA TYDMSGRVLL FFFNTTGRLI SDFQNFRITL
TIKALVEYRN DKRYLKIYNL VPSLDLDRWI IWLDGLYKEN
TDVTIFMNKL FNENWVEFWN DLQPGLVKAF TNAFTVLLNR
VFDNVAYDDM FLPYVDIRMG S

P09859; Gastrin/cholecystokinin-like peptide precursor. Gallus gallus.
MKTKVFLGLI LSAAVTACLC RPAAKAPGGS HRPTSSLARR
DWPEPPSQEQ QQRFISRFLP HVFAELSDRK **GFVQGNGAVE**
ALHDHFYPDW MDFGRRSTED AADAA

(28) Q9W0X2; CG9358 OR BCDNA:RE09339
MKASLALVFC VCVGLAAAAP EKTYTNKYDS VNVDEVLGNN
RVLGNYLKCL MDKGPCTAEG RELKRLLPDA LHSDCSKCTE
VQRK**NSQKVI NYLRANKAGE WKLLLNKYDP QGIYRAKHEG**
H

P01237; Prolactin precursor(PRL). Rattus norvegicus.
MNSQVSARKA GTLLLLMMSN LLFCQNVQTL PVCSGGDCQT
PLPELFDRVV MLSHYIHTLY TDMFIEFDKQ YVQDREFIAK
AINDCPTSSL ATPEDKEQAQ KVPPEVLLNL ILSLVHSWND
PLFQLITGLG GIHEAPDAII SRAKEIEEQN KRLLEGIEKI
ISQAYPEAKG NEIYLVWSQL PSLQGVDEES KDLAFYNNIR
CLRR**DSHKVD NYLKFLRCQI VHKNNC**

Table 3

Species	Peptide name	Sequence	Ref.
Mus musculus	γ-MSH	**YVMGHFRWDRFamide**	(55)
Rana catesbeiana	γ-MSH	YVMSHFRWNKFamide	(56)
Xenopus laevis (African clawed frog	γ-MSH	YVMTHFRWNKFamide	(57)
Oncorhynchus keta (chum salmon)	γ-MSH	HSYSMEHFRWamide	(58)
Theromyzon tessulatum	γ-MSH	YVMGHFRWDKF	(59)
Drosophila melanogaster	γ-MSH	WKILTAGSHFRWL	This study

Table 4

Species	**Peptide name**	**sequence**	**Ref.**
Orconectes limosus	Orcokinin [V^{13}]-Orcokinin	**NFDEIDRSGFGFN** NFDEIDRSGFGFNV	(33)
Carcinus maenes	[S^{9}]-Orcokinin [V^{13}]-Orcokinin [A^{13}]-Orcokinin	NFDEIDRSSFGFN NFDEIDRSGFGFNV NFDEIDRSGFGFNA	(60)
Procambarus clarkii	Orcokinin [V^{13}]-Orcokinin [A^{13}]-Orcokinin [T^{8}]-[H^{13}]-Orcokinin	NFDEIDRSSFGFN NFDEIDRSGFGFNV NFDEIDRSGFGFNA NFDEIDRTGFGFNH	(61)
Cherax destructor	Orcokinin [V^{13}]-Orcokinin [A^{8}]-[A^{13}]-Orcokinin [T^{8}]-[H^{13}]-Orcokinin	NFDEIDRSSFGFN NFDEIDRSGFGFNV NFDEIDRAGFGFNA NFDEIDRTGFGFNH	(62)
Cancer borealis	[V^{13}]-Orcokinin [A^{13}]-Orcokinin [S^{9}]-[V^{13}]-Orcokinin NFDEIDRSGFA	NFDEIDRSGFGFNV NFDEIDRSGFGFNA NFDEIDRSSFGFNV NFDEIDRSGFA	(34)
Drosophila melanogaster	Drm Orcokinin	NFDEIDKASASFSILN QLV	This study

Table 5

Name	Sequence	Ref
IM1	GNVIINGDCRVCNVHGamide	(15)
IM2	GNVVINGDCKYCNVHGamide	
IM3	GNVIINGDCRVCNVRA	
IM4	GTVLIQTDNTQYIRT	
CG15065	GNVIINGDCRHCNVRGamide	This study
CG5791	GNTIVNGRCQHCNVDPY	This study

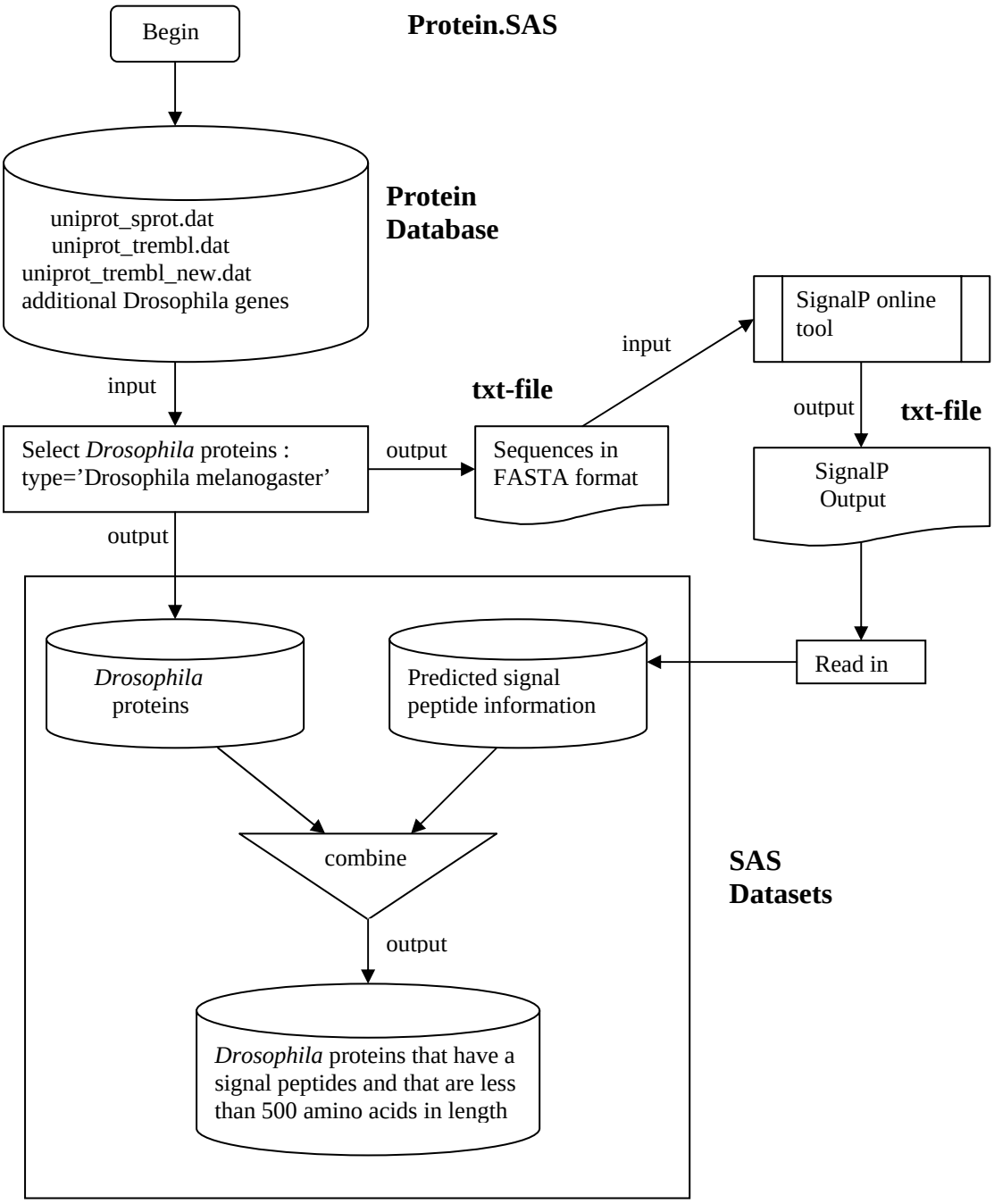
Begin
Protein.SAS
uniprot_sprot.dat
uniprot_trembl.dat
uniprot_trembl_new.dat
additional Drosophila genes
Protein Database
input
Select *Drosophila* proteins :
type='Drosophila melanogaster'
output
txt-file
Sequences in FASTA format
input
SignalP online tool
output
txt-file
SignalP Output
Read in
output
Drosophila proteins
Predicted signal peptide information
combine
output
Drosophila proteins that have a signal peptides and that are less than 500 amino acids in length
SAS Datasets

Fig1

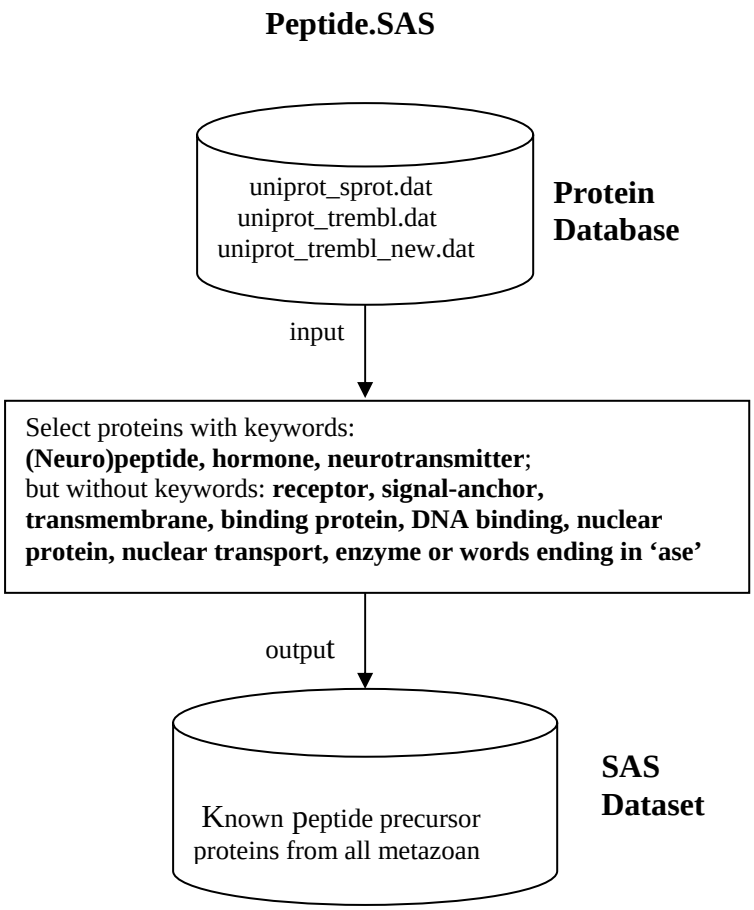
Peptide.SAS
uniprot_sprot.dat
uniprot_trembl.dat
uniprot_trembl_new.dat
Protein Database
input
Select proteins with keywords:
(Neuro)peptide, hormone, neurotransmitter;
but without keywords: receptor, signal-anchor, transmembrane, binding protein, DNA binding, nuclear protein, nuclear transport, enzyme or words ending in 'ase'
output
Known peptide precursor proteins from all metazoan
SAS Dataset

Fig 2

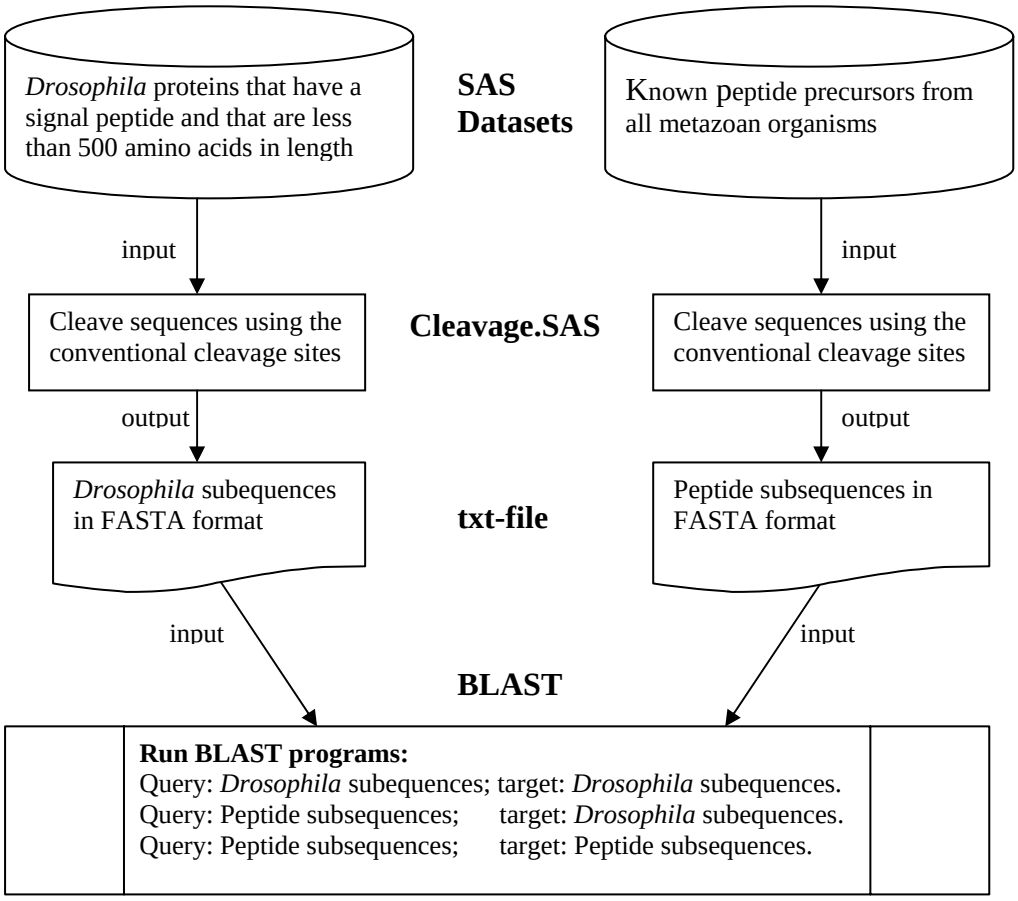
Drosophila proteins that have a signal peptide and that are less than 500 amino acids in length
SAS Datasets
Known peptide precursors from all metazoan organisms
input
input
Cleave sequences using the conventional cleavage sites
Cleavage.SAS
Cleave sequences using the conventional cleavage sites
output
output
Drosophila subequences in FASTA format
txt-file
Peptide subsequences in FASTA format
input
input
BLAST
Run BLAST programs:
Query: *Drosophila* subequences; target: *Drosophila* subequences.
Query: Peptide subsequences; target: *Drosophila* subequences.
Query: Peptide subsequences; target: Peptide subsequences.

Fig 3

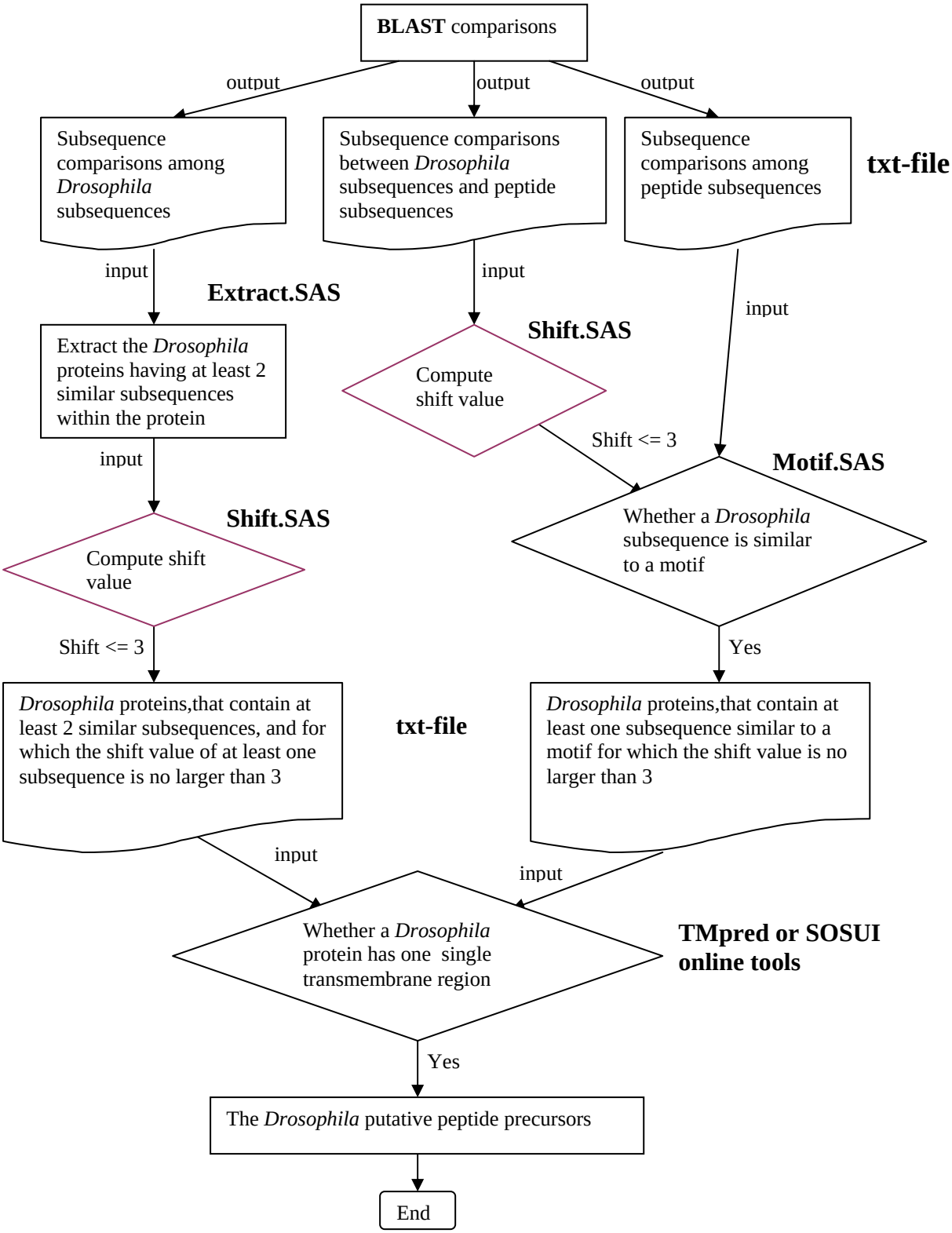
BLAST comparisons
output
output
output
Subsequence comparisons among Drosophila subsequences
Subsequence comparisons between Drosophila subsequences and peptide subsequences
Subsequence comparisons among peptide subsequences
txt-file
input
Extract.SAS
input
Shift.SAS
input
Extract the Drosophila proteins having at least 2 similar subsequences within the protein
Compute shift value
Shift <= 3
Motif.SAS
input
Shift.SAS
Compute shift value
Whether a Drosophila subsequence is similar to a motif
Shift <= 3
Yes
Drosophila proteins,that contain at least 2 similar subsequences, and for which the shift value of at least one subsequence is no larger than 3
txt-file
Drosophila proteins,that contain at least one subsequence similar to a motif for which the shift value is no larger than 3
input
input
Whether a Drosophila protein has one single transmembrane region
TMpred or SOSUI online tools
Yes
The Drosophila putative peptide precursors
End

Fig 4

```
Query=gnl|PID|Q9V808_6
(16 letters)

>gnl|PID|Q9V808_9
Length = 16

Score = 41.4 bits (90), Expect = 2E-6
Identities = 13/15 (86%), Positives = 13/15 (86%)

Query: 1  IPYEVKVDVPQPYIV 15
          IPYEVKV   VPQPY  V
Sbjct: 1  IPYEVKVPVPQPYEV 15

>gnl|PID|Q9V808_8
Length = 26

Score = 38 bits (82), Expect = 0.00003
Identities = 12/14 (85%), Positives = 12/14 (85%)

Query: 2  PYEVKVDVPQPYIV 15
          PYEVKV   VPQPY  V
Sbjct: 12 PYEVKVPVPQPYEV 25
```

The protein sequence 'Q9V808' (The similar subsequences are in bold. The match (similar amino acid sequence tags) is underlined.):

Q9V808; CG6564 protein (CG30101 OR CG6564 OR CG15901).*Drosophila melanogaster.*
MRMFVLPCLA VCVALAHCGG AVEDKKAEGD GKTVEKRGLH LGDYHHYQPH HEHIKTVTIE KKIPVPYTVT
KHVPYTVEKK **IPYEVKVDVP QPYIVE**KKVP VHVKEYVKVP VHVPKPYEVI KK**IPYEVKVP VDKPYEVKVP**
VPQPYEVIKK **IPYEVKVPVP QPYEVI**KKVP HEVKVEVPVP KPYEVIKKVP YEVKYEVEKP YDVEVPKPYD
VEVEKPYTVV VEKKVPYEVK VPVDKPYKVE VEKPYPVHVK VPVPQPYTVE KKVPYTVEKP VPYEVKVPIE
KPIPVYTEVK VPIHKEIPVP EKYHVEVPIF KHHQEDHHDY HSHGHGHY

Fig 5

The comparison between P11885_2 and Q8MS86_2:

```
Query=P11885_2
(11 letters)

>Q8MS86_2
Length = 13

Score = 22.3 bits (45), Expect = 1.1
Identities = 5/5 (100%), Positives = 5/5 (100%)

Query:   4 SHFRW 8
           SHFRW
Sbjct:   8 SHFRW 12
```

The comparison between P11885_2 and P06298_8:

```
Query=P11885_2
(11 letters)


>P06298_8
Length = 17

Score = 22.7 bits (46), Expect = 0.32
Identities = 6/8 (75%), Positives = 6/8 (75%)

Query:   1 YVMSHFRW 8
           Y M HFRW
Sbjct:   4 YRMHHFRW 11
```

The protein sequence 'Q8MS86', 'P11885' and 'P06298'(The similar subsequences are in bold. The match is underlined):

Q8MS86; LP04693p(BCDNA:LP04693).*Drosophila melanogaster*.
MSCSAWTQTP THTHKHRAIQ IVTIISVLII ECSALVACSL TPTSSLPALH RR**WKILTAGS HFRWL**

P11885;Corticotropin-lipotropin precursor. *Rana catesbeiana*.
MLQPVWHACI LAILGVFIFH VGEVRSQCWE SNKCTDLSSE DGILECIKAC KMDLSAESPV FPGNGHIQPL
SENIRK**YVMS HFRWNKF**GRR NSTSNDNNNN NGGYKREDIA NYPILNLFLG SDNQNTQEGI MEDDALDRQD
SKRSYSMEHF RWGKPVGKKR RPIKVFPTDA EEESSESFPI ELRRELSLEF DYPDTNSEEE LDNGELLEGP
VKKGRKYKMH HFRWEGPPKD KRYGGFMTPE RSQTPLMTLF KNAIIKNAHK KGQ

P06298;POMCA.Corticotropin-lipotropin A precursor.*Xenopus laevis.*
MFRPLWGCFL AILGICIFHI GEVQSQCWES SRCADLSSED GVLECIKACK TDLSAESPVF PGNGHLQPLS
ESIRKYVMTH FRWNKFGRRN STGNDGSNTG YKREDISSYP VFSLFPLSDQ NAPGDNMEEE PLDRQENKRA
YSMEHFRWGK PVGRKRRPIK VYPNGVEEES AESYPMELRR ELSLELDYPE IDLDEDIEDN EVKSALTKK**N**
GNYRMHHFRW GSPPKDKRYG GFMTPERSQT PLMTLFKNAI IKNSHKKGQ

Fig 6

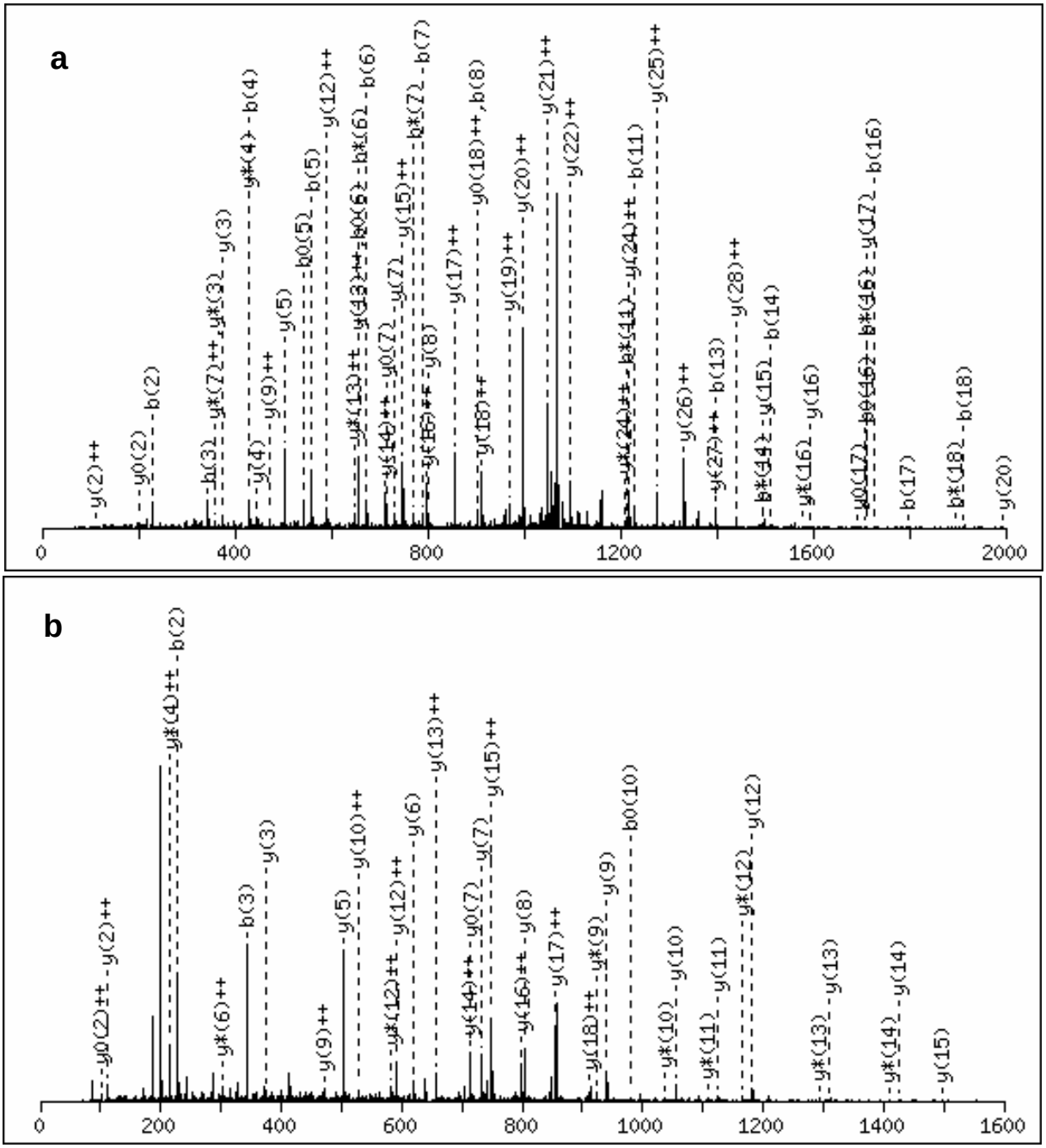
a
0
400
800
1200
1600
2000
b
0
200
400
600
800
1000
1200
1400
1600

Fig 7

(32) Q9V5V1; CG7738.

MRLTLLALIG VLCLACAYA***L*** ***DDSENNDQVV***
GLLDVADQGA ***NHANDGAREA*** RQLGGWGGGW
GGRGGWGGRG GWGGRGGWGG RGGWGGGWGG
RGGWGGRGGG WYGR

Fig 8