THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 43, Issue of October 25, pp. 40368–40374, 2002
Printed in U.S.A.

Peptidomics of the Larval *Drosophila melanogaster* Central Nervous System*

Received for publication, June 24, 2002, and in revised form, August 7, 2002
Published, JBC Papers in Press, August 8, 2002, DOI 10.1074/jbc.M206257200

Geert Baggerman, Anja Cerstiaens, Arnold De Loof, and Liliane Schoofs‡

From the Laboratory of Developmental Physiology and Molecular Biology, Katholieke Universiteit Leuven, Naamsestraat 59, B-3000 Leuven, Belgium

Neuropeptides regulate most, if not all, biological processes in the animal kingdom, but only seven have been isolated and sequenced from *Drosophila melanogaster*. In analogy with the proteomics technology, where all proteins expressed in a cell or tissue are analyzed, the peptidomics approach aims at the simultaneous identification of the whole peptidome of a cell or tissue, *i.e.* all expressed peptides with their posttranslational modifications. Using nanoscale liquid chromatography combined with tandem mass spectrometry and data base mining, we analyzed the peptidome of the larval *Drosophila* central nervous system at the amino acid sequence level. We were able to provide biochemical evidence for the presence of 28 neuropeptides using an extract of only 50 larval *Drosophila* central nervous systems. Eighteen of these peptides are encoded in previously cloned or annotated precursor genes, although not all of them were predicted correctly. Eleven of these peptides were never purified before. Eight other peptides are entirely novel and are encoded in five different, not yet annotated genes. This neuropeptide expression profiling study also opens perspectives for other eukaryotic model systems, for which genome projects are completed or in progress.

The most common approach in proteomic studies is to separate and visualize as many proteins as possible of an organism, tissue, or cell, by two-dimensional electrophoresis and to subsequently identify differentially expressed proteins by mass spectrometric techniques. One of the major constraints of this technology is that proteins of a molecular mass lower than 10 kDa are generally not retained and overlooked in most of the proteomic studies. Nevertheless this mass region contains a group of very important proteins, the peptide hormones and neurotransmitters.

To date, many neuropeptides have been purified from vertebrate and invertebrate sources. Peptide physiology in the pregenomic era (before the realization of the genome projects) was time-consuming as a peptide present in tissue extracts had to be purified to homogeneity prior to sequencing, synthesis, and functional analysis.

Drosophila is an outstanding model system for the molecular genetics and developmental biology in higher eukaryotes. Yet, its molecular endocrinology is poorly documented because only seven neuropeptides have been identified by traditional purification. Recently, a milestone was reached with the completion of the *Drosophila* genome project (1). Through the BLAST program, one can screen the genome of an organism for candidate-neuropeptides based on sequence homology with known neuropeptides from other organisms. In this way, 31 (neuro)peptide genes were found in *Drosophila melanogaster* (2–5). These findings, however, give no information on the temporal and spatial expression nor on the physiological relevance of these (neuro)peptide genes. Often, the precursors of biologically active peptides encode several peptides. In general, these peptides can be predicted from the precursor gene as it is assumed that proteolytic processing generally occurs at mono- and dibasic cleavage sites. However, not all of these possible cleavage sites are always used. Prediction of precursor-derived peptides is even more difficult when the existence of unconventional cleavage sites is taken into account (6). As a result, it is impossible to predict with certainty what the final processing products will be, based on the sequence of the precursor solely. This means that the structural characterization of the naturally occurring peptides remains necessary, not only to confirm the cleavage of the peptides from their precursor but also to identify posttranslational modifications, which cannot be predicted with certainty from the DNA sequence alone.

Along with the emergence of very sensitive mass spectrometry and the miniaturization of separation techniques, a new approach, designated as peptidomics in analogy with proteomics, enables us to analyze the entire peptidome of a tissue at the amino acid sequence level, starting from very little material. Peptidomics overcomes the drawback of the small body size of *Drosophila* that hampered traditional purification efforts.

In this study, we investigate the neuropeptide profile of the larval *Drosophila* central nervous system, using nanoflow capillary liquid chromatography-tandem mass spectrometry and the *Drosophila* genome data base as tools to find and identify neuropeptides

EXPERIMENTAL PROCEDURES

Animals—D. melanogaster are kept in 250-ml bottles and fed on a diet per bottle consisting of 70 ml of water, 17 g of sucrose, 0.45 g of yeast, 0.9 g of agar, 0.5 ml of 8% Nipagin, and 0.36 ml of propionic acid. Batches of 50 CNS[1] of "wandering" larvae are dissected and extracted in 20 μl of methanol/water/formic acid (90/9,9/0,1, v/v/v). The supernatants are filtered through Millipore spin down filters, dried, and stored at –20 °C.

Capillary LC-Tandem MS—Capillary LC-tandem MS experiments were conducted using an Ultimate high pressure LC pump, a column-switching device (Switchos) and a Famos autosampler (all LC Packings) coupled to a quadrupole time-of-flight mass spectrometer (Micromass).

* This project was sponsored by the Flemish Science Foundation (FWO) Grants G.0356.98, G.0187.00, and G.0175.02 and by the Katholieke Universiteit Leuven Grant GOA/2000/04. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 32-1632-4260; Fax: 32-1632-3912; E-mail: liliane.schoofs@bio.kuleuven.ac.be.

[1] The abbreviations used are: CNS, central nervous system; LC, liquid chromatography; MS, mass spectrometry; FARP, FMRF-amide-related peptide; NPF, neuropeptide F; AKH, adipokinetic hormones; DIM, *Drosophila* immune-induced.

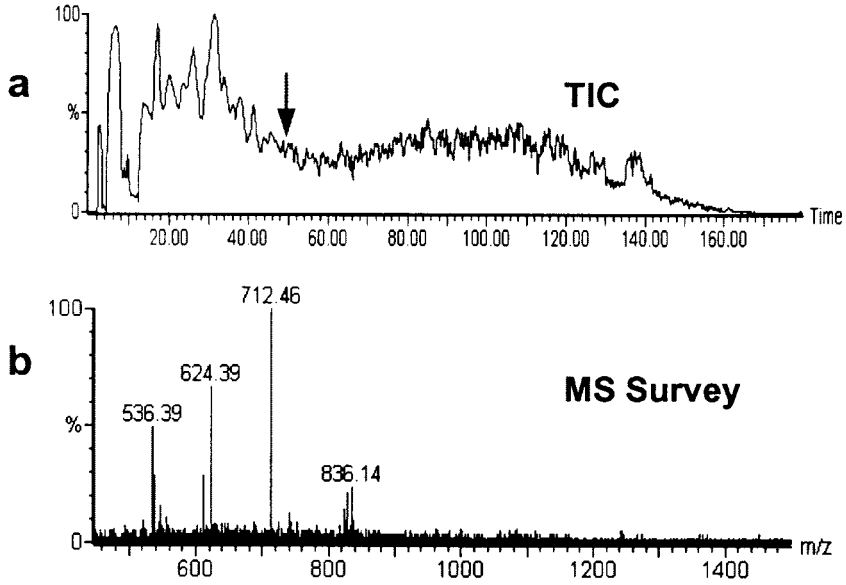

FIG. 1. *a*, total ion current chromatogram obtained by nanoLC-MS analysis of a methanolic extract of 50 larval *Drosophila* CNSs. *b*, example of one survey mass spectrum taken from the total ion current chromatogram (indicated by an *arrow*). Two double-charged ions from this spectrum were selected for fragmentation (624.39 *m/z* and 712.47 *m/z*). The resulting fragmentation spectra and their amino acid sequence annotation are shown in Fig. 2.

Chromatography was performed using a guard column (μ-guard column MGU-30 C18, LC-Packings) 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 MilliQ 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, 3 μm of 75 μm $\times$ 150 mm nano column (LC Packings). Separation was conducted using a linear gradient from 95% solvent A and 5% solvent B to 20% A and 80% B in 90 min, followed by a linear gradient from 20% A and 80% B to 50% A and 50% B in 60 min (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 is connected in series to the electrospray interface of the quadrupole time-of-flight device. The column eluent was directed through a metal-coated fused silica tip (Picotip type FS360-75-10 D, New Objective). Needle voltage was set at 1300 volts, cone voltage at 35 volts. Nitrogen was used as nebulizing gas. All double and triple-charged ions (typical for peptides, background ions have generally only one charge) of sufficient parent ion intensity (threshold was set at 15 counts/second) were automatically recognized by the charge state recognition software (Masslynx 3.5, Micromass) and selected for fragmentation as they eluted from the column. Argon was used as a collision gas; collision energy was set at 25–40 eV depending on the mass and charge state of the selected ion.

Peptide Identification—Three different successive approaches were used for peptide identification. 1) The fragmentation spectrum of an ion with an experimental mass corresponding to the calculated mass of a known or predicted *Drosophila* peptide was compared and matched with the theoretical fragmentation spectrum of that peptide obtained from the Biolynx software (Micromass). 2) Unknown peptides were identified by subjecting their fragmentation spectra to a Mascott search (www.matrixscience.com). This bioinformatic tool allows the identification of a peptide in a (genomic) data base by analyzing the fragmentation spectrum of the peptide and by using the results of this analysis to search the data base for a possible match. 3) Alternatively, the fragmentation spectrum was first analyzed manually, and the deduced amino acid sequence as well as the total peptide mass were used as queries in a Mascot search. Often only a partial amino acid sequence is obtained. Nevertheless, the Mascot program can be used to identify the gene from the genomic data base using the partial sequence and the parent mass as queries. All obtained fragmentation spectra were analyzed in this way.

Prediction of Signal Peptides—The programs TargetP (version 1.01) (www.cbs.dtu.dk/services/TargetP/) (7, 8) and SignalP (version 1.1), (www.cbs.dtu.dk/services/SignalP/) were used to predict the occurrence of a signal peptide and the cellular localization of each *Drosophila* neuropeptide precursor sequence identified (see Table II).

RESULTS

The nanoLC-tandem MS technology allows us to select and fragment the peptide ions as they elute from the column even when co-eluting with other peptides. As an example, Figs. 1 and 2 represent an example of the selection and sequencing of two co-eluting peptides. All multiple-charged ions (double, triple, or quadruple) were selected and fragmented in the same way.

In total we were able to sequence 18 peptides that are derived from 10 of the 31 annotated *Drosophila* (neuro)peptide precursors and eight novel peptides that are derived from five novel (not annotated) precursors and two immune-induced peptides, so far only encountered in adult hemolymph. The sequences are presented in Table I, and the nomenclature is derived from the flybase. The used programs predict an extracellular localization for all 15 peptide precursor proteins and yield a potential signal peptide, except for the leucokinin precursor.

FMRF-amide-related Peptides or FARPs—In *D. melanogaster*, as in other insects, a high number of neuropeptides display a C-terminal amino acid sequence, similar to FMRF-amide. In the fruit fly, these FARPs are encoded by four different genes. In our experiments, we found peptides originating from all four precursor proteins.

The *Fmrf* gene contains five copies of DPKQDFMRF-amide, as well as 10 additional possible amidated peptides (9). Three of the encoded peptides (DPKQDFMRF-amide, TPAEDFMRF-amide (2 copies/gene), and SDNFMRF-amide) have already been purified and sequenced from extracts of adult *Drosophila* (10). We were able to detect these Drm-FMRF-amide peptides in the larval CNS extracts along with another peptide (PDN-FMRF-amide) predicted from the same *Fmrf* gene.

The dromyosuppressin or *Dms* gene only encodes the myoinhibitory peptide dromyosuppressin or TDVDHVLRF-amide (10). In the LC-MS experiment we were able to detect and fully sequence dromyosuppressin (Fig. 2*a*) as one of the most prominent MS peaks.

The drosulfakinin or *Dsk* gene has been cloned (11) and predicts three putative peptides, Drm-SK-0, Drm-SK-1, and Drm-SK-2, annotated as drosulfakinins because of their sequence homology with other insect sulfakinin peptides and with the vertebrate hormones of the gastrin/cholecystokinin family. Similar to cholecystokinin in vertebrates, sulfakinins reduce food intake and stimulate muscle contractions in insects (12). An ion with a fragmentation spectrum corresponding Drm-SK-2 (GGDDQFDDY(SO3H)GHMRF-amide) with a sulfated tyrosine residue was detected.

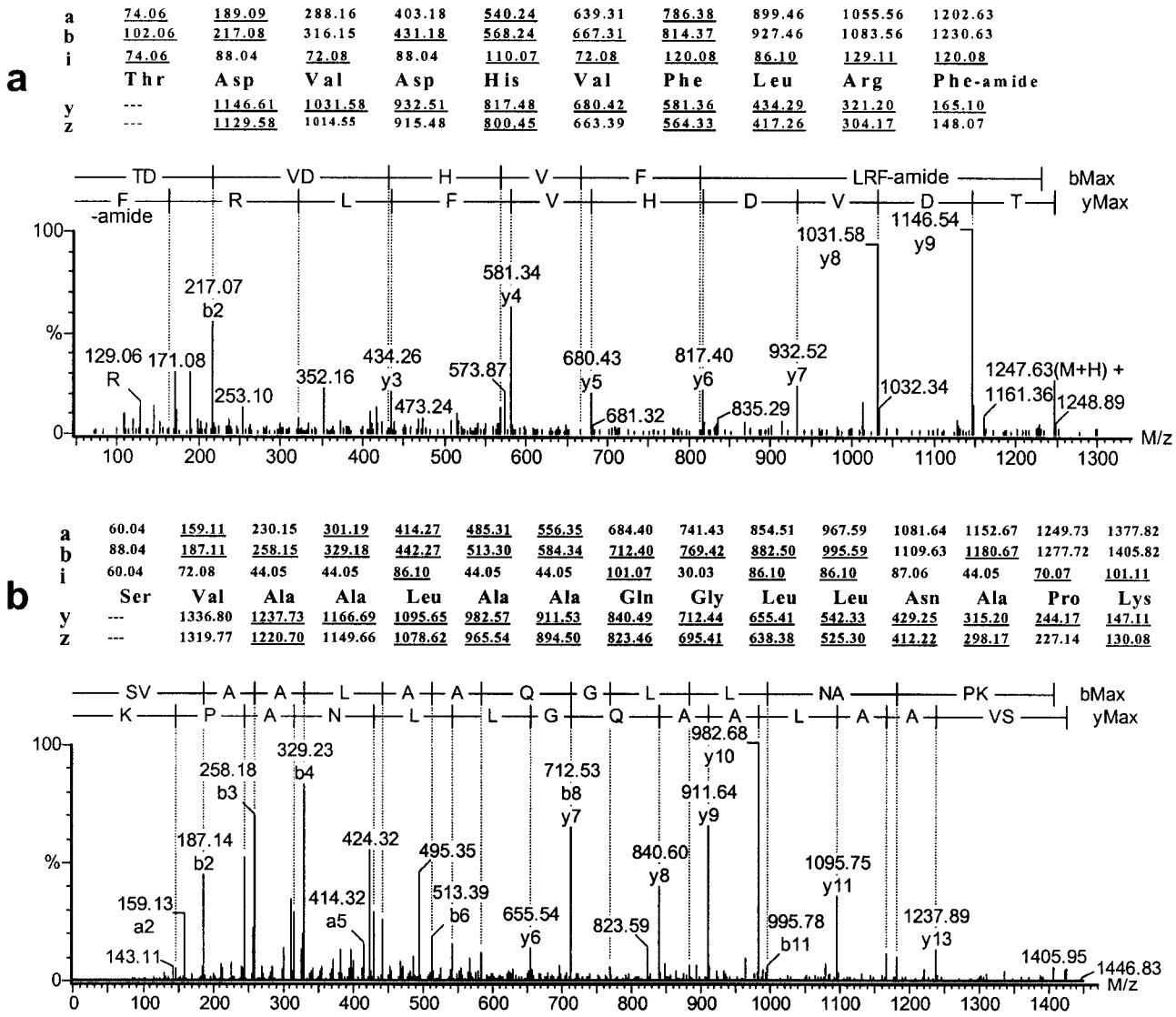

FIG. 2. *a*, fragmentation spectrum of the peptide at 624.39 *m/z* (Dromyosuppressin). a-type, b-type, y-type, z-type and immonium (*i*) fragment ions are indicated. The theoretical fragment ion masses found in the spectrum are underlined. "-amide" indicates a C-terminal amidation. *b*, fragmentation spectrum of the peptide at 712.47 *m/z*. The amino acid sequence was determined as SVAALAAQGLLNAPK annotated by a new peptide precursor annotated in the *Drosophila* genome data base as CG13061. b-type and y-type fragment ions are indicated in the spectrum.

The *sNPF* gene has not yet been cloned but was found by BLAST analysis (3). It encodes four putative peptides. Two of them display sequence similarities with two short NPFs from the Colorado potato beetle, *Leptinotarsa decemlineata*, and are annotated as Drm-sNPF1 (AQRSPSLRLRF-amide) and Drm-sNPF2 (WFGDVNKPIRSPSLRLRF-amide). None of these predicted peptides were detected in our mass spectral read-outs. Instead, other peptides derived from the same precursor were detected and sequenced. We annotated the first one, SDPDMLNSIVE, as Drm-sNPF-AP or *Drosophila*-short NPF-associated peptide, referring to its presence in the same precursor as Drm-sNPF. In its precursor Drm-sNPF-AP is flanked by two dibasic cleavage sites (Table II). It was identified in its non-oxidized form as well as with an oxidized methionine residue. It is unclear whether this is a natural form or whether the oxidation occurred during sample preparation. The second identified peptide originating from this precursor is WFGDVNKPI and corresponds to the first nine amino acids of a Drm-sNPF2. This seems to indicate that the arginine at position 10 acts as a monobasic cleavage site. Therefore we name this novel peptide Drm-sNPF-$2^{1–9}$.

Adipokinetic Hormone and Corazonin—The adipokinetic hormone (*Akh*) gene and the corazonin (*Crz*) gene display structural similarities suggesting that they have a common evolutionary origin. Adipokinetic hormones (AKH) regulate the release of lipid and carbohydrate energy substrates during flight. *Drosophila* AKH (13) is encoded by the *Akh* gene, which is structurally similar to the three AKH genes in the locust, *Locusta migratoria* (14). Although the level of AKH expression in the corpora cardiaca of locusts is extremely high (6, 15), we observed only a weak Drm-AKH signal. No other product from the AKH precursor was found. Corazonin is an insect neuropeptide having cardio-acceleratory as well as pigmentotropic activities (16). The *Crz* gene was cloned from a genomic library of *D. melanogaster*. In addition to a putative corazonin, it encodes for a 39-amino acid corazonin-precursor-related peptide (17). Our data unequivocally demonstrate the presence of corazonin in the larval CNS, as the fragmentation spectrum of the double-charged ion at 685.3 *m/z* corresponds with the amino acid sequence of corazonin. However, the peptide predicted as corazonin-precursor-related peptide (17) could not be detected.

TABLE I

Peptides identified in the larval Drosophila melanogaster CNS by nano-LC-tandem MS

Peptides previously purified and sequenced in *Drosophila* are in italic face. pQ, pyroglutamine residue.

Peptide name (BDGP annotation)	Peptides characterized in larval *Drosophila* brain by LC-MS/MS (Monoisotopic mass)	Ref.
	FaRP family	
FMRF-amide (CG2346)		37
Drm-FMRFa 1	*DPKQDFMRF-amide* (1111.51 Da)	
Drm-FMRFa 2	*TPAEDFMRF-amide* (914.41 Da)	
Drm-FMRFa 3	*SDNFMRF-amide* (924.43 Da)	
Drm-FMRFa 4	PDNFMRF-amide (1181.56 Da)	
Dromyosupressin (CG6440)		10
Drm-MS	*TDVDHVFLRF-amide* (1246.64 Da)	
drosulfakinin (CG18090)		11
Drm-SK-2	GGDDQFDDY(SO3)GHMRF-amide (1737.61 Da)	
Drm-sNPF (CG13968)		3
Drm-sNPF 2^{1-10}	WFGDVNQKPI (1202.61 Da)	
Drm-sNPF-AP 1	SDPDMLNSIVE (1218.45 Da)	
Drm-sNPF-AP 1	SDPDM(ox)LNSIVE (1234.53 Da)	
	Adipokinetic hormone family	
AKH (CG1171)		13
Drm-AKH	*pQLTFSPDW-amide* (974.45 Da)	
corazonin (CG3302)		17
Drm-COR	pQTFQYSRGWTN-amide (1368.62 Da)	
	Allatostatic peptides	
Allatostatin (CG13633)		18
Drm-AST-2^{1-11}	AYMYTNGGPGM (1160.46 Da)	
Drm-AST-3	SRPYSFGL-amide (924.48 Da)	
Drm-AST-4	TTRPQPFNFGL-amide (1275.67 Da)	
MIP-like peptide (CG6456)		19
Drm-MIP-2	AWKSMNVAW-amide (1090.54 Da)	
Drm-MIP-5	DQWQKLHGGW-amide (1252.61 Da)	
	Diuretic peptides	
CAP2b peptide (CG15520)		3
Drm-CAP-1	GANMGLYAFPRV-amide (1293.67 Da)	
Drm-CAP-2	ASGLVAFPRV-amide (1014.6 Da)	
DroLeucokinin (CG13480)		27
DLK	*NSVVLGKKQRFHSWG-amide* (1740.95 Da)	
	Immune-induced peptides	
Immune-induced protein 2 (IM2)		29
DIM2	*GNVVINGDCKYCVNG-amide* (1689.77 Da)	
Immune-induced protein 4 (CG15231)		29
DIM4	*GTVLIQTDNTQYIRT-amide* (1720.91 Da)	
	Newly characterized peptides	
Neuropeptide like precursor 1 (CG 3441)		
MTYamide peptide	YIGSLARAGGLMTY-amide (1470.76 Da)	
IPNamide peptide	NVGTLARDFQLPIPN-amide (1652.89 Da)	
APK peptide	SVAALAAQGLLNAPK (1422.81 Da)	
Hugin (CG6371)		
Drm-MT2	SVPFKPRL-amide (841.58 Da)	
Neuropeptide like precursor 2 (CG11051)		
NEF peptide	TKAQGDFNEF (1155.52 Da)	
Neuropeptide like precursor 3 (CG13061)		
SHA peptide	VVSVVPGAISHA (1134.65 Da)	
VVIamide peptide	SVHGLGPVVI-amide (975.58 Da)	
Neuropeptide like precursor 4 (CG15361)		
YSY peptide	pQYYYGASPYAYSGGYYDSPYSY (2567.01 Da)	

Allatostatic Peptides—Allatostatins are insect neuropeptides that block the synthesis of juvenile hormone (a sesquiterpenoid) synthesis by the corpora allata, two endocrine organs near the insect brain. In *Drosophila*, three different genes (*Ast*, *Ast2*, and *MIP*) encode putative allatostatins. The *Ast* gene encodes four putative allatostatins (18). We found two of them in the larval CNS: SRPYSFGL-amide (drostatin 3), which was recently isolated as a ligand for an orphan receptor (19), and TTRPQPFNFGL-amide (drostatin 4), which was predicted from the precursor but not yet isolated. In addition, we identified a third peptide, AYMYTNGGPGM, which is contained within the predicted drostatin 2 or Drm-AST-2, which itself was not detected. AYMYTNGGPGM is separated by a dibasic site (KR) from the remaining part of Drm-AST-2 (Table II) suggesting that this typical cleavage site had been overlooked in the prediction study (3). AYMYTNGGPGM displays no sequence similarity with any other known peptide. The single peptide, predicted from the *Ast2* gene was not detected in our MS read-outs.

The third allatostin gene, called *MIP* gene has recently been cloned (19) and encodes five putative peptides with sequence similarities to *Locusta*-MIP (20). MIPs are neuropeptides with myoinhibiting and allatostatic activities. No *Drosophila*-MIP could as yet be isolated. We were able to detect and sequence 2 of the five predicted peptides in the larval CNS extract. Both peptides, AWKSMNVAW-amide (predicted as Drm-MIP-2) and

TABLE II

Precursors from which peptides were identified in larval Drosophila brain by LC-MS/MS

Consensus cleavage sites are underlined, identified peptides are in bold face, signal peptides, as predicted by signal P, are in italic face.

FMRFamide precursor (CG2346/*fmrf* gene)

MGIALMFLLA LYQMQSAIHS EIIDTPNYAG NSLQDADSEV
SPPQDNDLVD ALLGNDQTER AELEFRHPIS VIGIDYSKNV
VLHFQKHGRK PRYKYDPELE AKRRSVQDNF MHFGKRQAEQ
LPPEGSYAES DELEGMAKRA AMDRYGRDPK **QDFMRF**GRDP
KQDFMRFGRD **PKQDFMRF**GR **DPKQDFMRF**G **RDPKQDFMRF**
GR**TPAEDFMR FGRTPAEDFM RFGRSDNFMR FGRSPHEELR**
SPKQDFMRFG *RPDNFMRF*GR SAPQDFVRSG KMDSNFIRFG
KSLKPAAPES KPVKSNQGNP GERSPVDKAM TELFKKQELQ
DQQVKNGAQA TTTQDGSVEQ DQFFGQ

Dromyosuppressin precursor (CG6440/*Dms* gene)

MSFAQFFVAC CLAIVLLAVS NTRAAVQGPP LCQSGIVEEM
PPHIRKVCQA LENSDQLTSA LKSYINNEAS ALVANSDDLL
KNYNKR**TDVD HVFLRF**GKRR

Drosulfakinin precursor (CG18090/*Dsk* gene)

MPLWALAFCF LVVLPIPAQT TSLQNAKDDR RLQELESKIG
GEIDQPIANL VGPSFSLFGD RRNQKTMSFG RRVPLISRPI
IPIELDLLMD NDDERTKAKR FDDYGHMRFG KR**GGDDQFDD**
YGHMRFGR

Short NPFs precursor (CG13968/*sNPF* gene)

MFHLKRELSQ GCALALICLV SLQMQQPAQA EVSSAQGEHL
VQPPPEKQSS KDSFLGTPLS NLYDNLLQRE YAGPVVFPNH
QVERKAQRSP SLRLRFGR**SD PDMLNSIVE**K R**WFGDVNQKP**
IRSPSLRLRF GRRDPSLPQM RRTAYDDLLE RELTLNSQQQ
QQQLGTEPDS DLGADYDGLY ERVVRKPQRL RWGRSVPQFE
ANNADNEQIE RSQWYNSLLN SDKMRRMLVA LQQQYEIPEN
VASYANDEDT DTDLNNDTSE FQREVRKPMR LRWGRSTGKA
PSEQKHTPEE TSSIPPKTQN

AKH precursor (CG1171/*Akh* gene)

MNPKSEVLIA VLFMLLACQC **QLTFSPDWGK** RSVGGAGPGT
FFETQQGNCK TSNEMLLEIF RFVQSQAQLF LDCKHRE

Corazonin precursor (CG3302/*Crz* gene)

MLRLLLLPLF LFTLSMCMGQ **TFQYSRGWTN** GKRSFNAASP
LLANGHLHRA SELGLTDLYD LQDWSSDRRL ERCLSQLQRS
LIARNCVPGS DFNANRVDPD PENSAHPRLS NSNGEVLYSS
ANIPNRHRSN ELLEELSAAG GASAEPNVFG KH

Allatostatin precursor (CG13633/*Ast* gene)

*MNSLHAHLLL LAVCCVGYIA SS*PVIGQDQR SGDSDADVLL
AADEMADNGG DNIDKRVERY AFGLGRR**AYM YTNGGPGM**KR
LPVYNFGLGK R**SRPYSFGLG** KRSDYDYDQD NEIDYRVPPA
NYLAAERAVR PGRQNKR**TTR PQPFNFGL**GR R

Myoinhibiting peptide precursor (MIP) (CG6456/*MIP* gene)

MAHTKTRRTY GFLMVLLILG SACGNLVASG SAGSPPSNEP
GGGGLSEQVV LDQLSESDLY GNNKRAWQSL QSSWGKRSSS
GDVSDPDIYT GHFVPLVITD GTNTIDWDTF ERLASGQSAQ
QQQQQPLQQQ SQSGEDFDDL AGEPDVEKR**A WKSMNVAW**GK
RRQAQGWNKF RGAWGKREPT WNNLKGMWGK R**DQWQKLHGG**
WGKRSQLPSN

CAP2b myotropin precursor (CG15520/*Cap2b* gene)

MKSMLVHIVL VIFIIAEFST AETDHDKNRR **GANMGLYAFP**
RVGRSDPSLA NSLRDGLEAG VLDGIYGDAS QEDYNEADFQ
KK**ASGLVAFP RV**GRGDAELR KWAHLLALQQ VLDKRTGPSA
SSGLWFGPRL GKRSVDAKSF ADISKGQKEL N

Droleucokinin precursor (CG13480/*Leucokinin*)

MVLLAFGRQV YGASLVPAPI SEQDPELATC ELQLSKYRRF
ILQAILSFED VCDAYSSRPG GQDSDSEGWP FRHYAPPPTS
QRGEIWAFFR LLMAQFGDKE FSPIIRDVIE RCRIKSQLQR
DEKR**NSVVLG KKQRFHSWG**G KRSPEPPILP DY

Immune induced peptide 2 (IM2)

MKFFSVVTVF VFGLLALAN AVPLSPDP**GNV VINGDCKYCN**
VHGGK

Immune induced peptide 4 (CG15231)

MKFFQAAALL LAMFAALANA EPVPQP**GTVL IQTDNTQYIR T**G

Newly identified peptide precursors

Neuropeptide like precursor 1 (CG 3441)

*MQAVLQSAHS SRRLMLLLSM LLNAAIQ*PRS IIVSATDDVA
NVSPCEMESL INQLMSPSPE YQLHASALRN QLKNLLRERQ
LAVGEEQPLG EYPDYLEEDK R**SVAALAAQG LLNAP**KRSLA
TAKNGQLPTA EPGEDYGDAD SGEPSEQKR**Y IGSLARAGGL**
MTYGKR**NVGT LARDFQLPIP N**GKRNIATMA RLQSAPSTHR
DPKRNVAAVA RYNSQHGHIQ RAGAEKRNLG ALKSSPVHGV
QKREDEELLP AAAPDYADPM QSYWWYPSYA GYADLDWNDY
RRAEKRFLDT SKDPELFGIH GNDATTAEPA DEAYMESDAE
AGSEQLPSPQ KRHIGAVYRS GFLPSYRYLR SPGGSSGFGG
AGGRFSRSGR DARQFV

(Accession number AE003466)

HUGIN (CG6371)

*MCGPSYCTLL LIAASCYILV CSHA*KSLQGT SKLDLGNHIS
AGSARGSLSP ASPALSEARQ KRAMGDYKEL TDIIDELEEN
SLAQKASATM QVAAMPPQGQ EFDLDTMPPL TYYLLLQKLR
QLQSNGEPAY RVRTPRLGRS IDSWRLLDAE GATGMAGGEE
AIGGQFMQRM VKK**SVPFKPR LG**KRAQVCGG D

(Accession number AJ133105)

Neuropeptide like precursor 2 (CG11051)

MAKLAICILV FALFALALSA RVPREESNPA QEFL**TKAQGD**
FNEFIEKLKA LDAKKVEGLF KDGLNTVQEG LQKLNETFLQ
APAAST

(Accession number AE003538)

Neuropeptide like precursor 3 (CG13061)

MFKLCVFVAL LSLAAAAPAP APAPAPAPGL IGPGIVAPGI
WGPTVVGSPL LAPQ**VVSVVP GAISHA**AITQ VHPSPLLIK**S**
VHGLGPVVIG

(Accession number AE003527)

Neuropeptide like precursor 4 (CG15361)

MFKLLVVVFA ALFAAALAVP APVARANPAP IPIASPEPA**Q**
YYYGASPYAY SGGYYDSPYS YYG

(Accession number AE003584)

DQWQKLHGGW-amide (predicted as Drm-MIP 5), are amidated, which is in agreement with the C-terminal glycine residues in the precursor (Table II).

Peptides Derived from the capability or capa Gene—Cardio acceleratory peptide 2b (CAP2b), first isolated from the tobacco hawkmoth *Manduca sexta* (21), displays cardio-acceleratory properties and stimulates Malpighian tubule secretion (22). Querying the *Drosophila* genome for a gene that encodes for CAP2b-like peptides revealed a peptide precursor gene, which has been cloned (38) and which encodes two putative CAP2b-like peptides and one myotropin/pyrokinin peptide (Drm-MT) having a C-terminal FXPRL-amide motif. Myotropins are a family of peptides with members in several insect orders (23, 24) and in crustaceans (25). They have been found to regulate

a large number of processes in insects such as oviduct and hindgut contractions, sex pheromone synthesis, egg diapause, melanization, etc. (26). In the *Drosophila* CNS we identified two predicted CAP2b-like peptides, namely GANMGLYAFPRV-amide (Drm-CAP-1) and ASGLVAFPRV-amide (Drm-CAP-2). However, we did not find the predicted Drm-MT.

Leucokinin—We detected in the fruit fly CNS extract a triple-charged peptide ion at 581.3 *m/z* , whose fragmentation spectrum corresponds to the sequence of *Drosophila* leucokinin (NSVVLGKKQRFHSWG-amide). This peptide stimulates secretion in *D. melanogaster* Malpighian tubules (27).

Newly Identified Peptides and Their Precursors (Table II)—Eight entirely novel peptides were discovered that originate from five not yet annotated genes. Three of the newly discovered peptides originate from the same gene, annotated in the flybase as CG3441. The derived peptide precursor has a structure typical for a neuropeptide precursor (Table II) as the three novel peptides are flanked by consensus processing sites (KR and GKR). The novel peptides are YIGSLRAGGLMTY-amide, NVGTLARDFQLPIPN-amide and SVAALAAQGLLNAPK (Fig. 2*b*). Because nothing is as yet known about their possible functions, we named them MTY-amide, IPN-amide, and APK-amide, respectively, according to their final three C-terminal amino acid residues. The precursor was designated as neuropeptide-like precursor 1 (NPLP1) and it encodes nine other putative peptides, all flanked by dibasic cleavage sites. None of these peptides shares sequence homology to a known invertebrate or vertebrate neuropeptide.

A fourth novel *Drosophila* peptide (SVPFKPRL-amide, designated as Drm-MT-2), also detected in this study, has an F*X*PRL-amide-like C-terminal amino acid sequence, typical for myotropin/pyrokinin peptides. It is not encoded by the myotropin/CAP2b (*capability*) gene, but by a gene annotated as HUGIN/CG6371 in which it is flanked by consensus processing sites. SVPFKPRL-amide seems to be the only peptide encoded by this precursor gene.

The final three novel peptide precursors that we identified (CG13061, CG11051, and CG15361) were designated as neuropeptide-like precursor 2, 3, and 4, respectively. Within these respective precursors, the novel peptides, TKAQGDFNEF (1155.52 Da), SVHGLGPVVI-amide (975.58 Da) and VVSVVPGAISH (1134.65 Da) designated as NEF, VVI-amide, and ISH, are not flanked by typical dibasic cleavage sites. Nevertheless, one of these peptides SVHGLGPVVI-amide is indeed amidated and is, in its precursor, preceded by a lysine residue, which is a possible monobasic cleavage site. Another peptide pQYYYGASPYAYSGGYYDSPYDY (2567.0 Da, designated YDY), encoded in a genomic sequence annotated as CG15361, carries a pyroglutamic acid residue at its N terminus. The enzymes responsible for these modifications are found in the secretory pathway suggesting that these peptides are indeed secreted (28). Furthermore, TargetP queries also indicate that these peptides may have an extracellular localization (Table II).

Infection-induced Peptides—In adult *Drosophila*, immune-induced (DIM) peptides appear in the hemolymph upon infection (29). Our nano-LC-MS-MS experiment yielded the following amino acid sequences: GNVV(I/L)NGDC(K/Q)YCNG-amide and GTV(L/I)(L/I)(L/I)(Q/K)TDNT(Q/K)YIRT-amide. The non-amidated form of the first peptide was also found. These sequences correspond to those of *Drosophila* immune-induced protein 2 and 4 (DIM2, GNVVINGDCKYCNG-amide and DIM4, GTVLIIQTDNTQYIRT-amide). This is the first report on the presence of DIM-peptides in the larval CNS. Because our larvae were not infected prior to dissection it might be that these DIMs are produced in the CNS and released after infection, acting as modulators in the immune response.

DISCUSSION

With its complete genome sequence available, *Drosophila* becomes a model insect for peptidomic studies. From an extract of 50 larval CNSs only, we were able to fully identify 28 neuropeptides, eight of which are derived from not yet annotated genes. Only seven of these peptides had been isolated and sequenced from *Drosophila* before using traditional purification and amino acid sequencing (Edman chemistry) techniques, usually starting from several 100,000s *Drosophila* individuals for extraction. The present study illustrates the remarkable improvement in sensitivity provided by the combination of nanoscale LC, electrospray time-of-flight mass spectrometry. Indeed, we proved that the peptidomics approach can be used to identify peptides that cannot be mined from the genome based on sequence similarities with peptides from other animal sources. The presently identified eight new peptides, encoded by five not yet annotated genes, would have remained unidentified in the "pregenomic era". These results suggest that substantially more than the 24 predicted neuropeptide precursors are encoded by the *Drosophila* genome. This is in accordance with the bioinformatic analysis of the *Caenorhabditis elegans* genome, predicting the presence of 92 peptide genes (30).

However, several predicted neuropeptides, such as tachykinins, crustacean cardio-active peptide, ecdysis-triggering hormones, allatostatin C, eclosion-triggering hormone, and pigment dispersing hormone (31–33) could as yet not be detected in the CNS extract of the fruit fly. Predicted peptides with a higher molecular mass (>3.5 kDa), such as (long) Drm-NPF, Drm-ion transport peptide-like peptide, corticotropin-releasing factor-like diuretic hormone, Drm cAMP-generating peptide, calcitonin- and insulin-related peptides could not be detected either. Their identification is more problematic since the molecular mass limit for fragmentation in the collision cell of the quadrupole time-of-flight mass spectrometer is about 3.5–4 kDa. Although it is possible to identify these peptides based on their molecular mass solely (5) we did not find the corresponding masses. The absence of these peptides may be due to several causes. First off all, their concentration in the CNS may be below the sensitivity of the instrumental setup we used. Second, not all peptides are extracted or ionize with the same efficiency. Consequently, instrument sensitivity varies between peptides depending on their amino acid composition. Third, many peptides were only predicted from the genomic data base, which is no guarantee that the coding genes are expressed at all in the larval CNS. This is likely to be the explanation for the absence of ions corresponding to eclosion-triggering hormones, which in other insects are produced in epitracheal gland cells (Inka cells) (34). A fourth cause might be that the predicted peptide processing is incorrect or that some parts of the genome may not have been correctly sequenced. This might be the case for the predicted *Drosophila* diuretic hormone precursor (CG8348). The C-terminal part, in contrast to the N terminus of this putative diuretic hormone, cannot be nicely aligned with the C-terminal sequences of the other insect CRF-related peptides. Fifth, the expression of neuropeptide genes might be developmentally regulated. In this study we used larvae in the wandering stage, the stage just prior to pupariation, because these can be easily discerned from other larval stages. However, in this stage the larva undergoes many physiological changes, some of which are likely to be regulated by neuropeptides. Crustacean cardioactive peptide and eclosion hormone for instance, are known to be play a role in controlling ecdysis (35, 36). Release of these peptide hormones may result in their depletion from neurohemeal release sites in the CNS, hampering their identification.

Despite the sensitivity of the MS method, some peptides

most likely remained unidentified in this study, but it is difficult to determine how many. At least 100 parent ions of low intensity that were fragmented could not be sequenced because of the low quality and signal intensity of the obtained daughter ions, rendering identification through a Mascot search impossible. However, most of these parent ions may not be peptides. In most of these cases, it was impossible to determine whether the parent ion peak corresponded to a peptide or to another, non-peptidergic compound that might be present in the crude acid methanolic extract. Only in some cases, the fragmentation spectra (of 10–15 putative peptides) yielded a partial but inconclusive amino acid sequence (3–5 amino acids), too short to be useful in data base searches such as Mascot, which is written for proteomic studies. In the future, we intend to improve signal intensity by using 2- or *n*-dimensional LC-MS. This will allow the use of larger extracts (now limited to 50 central nervous systems of *Drosophila* larvae).

Historically, most endocrine factors have been defined based on changes in phenotype or in particular physiological processes as monitored by diverse bioassays. The peptidomics approach described herein will allow studying differential peptide expression of the entire peptidome in different physiological and developmental conditions. The traditional biochemical characterization of endocrine peptidic factors will not become completely obsolete, in particular for those peptides that resist identification by MS. For sure, the peptidomics technology will represent an opportunity to discover new peptide hormones and signaling molecules of the endocrine system, not only from *Drosophila* but from any metazoan of which the genome has been sequenced (*C. elegans*, man) or will be in the future.

REFERENCES

1. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M.,Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J.,Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D.,Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., and Venter, J. C. (2000) *Science* **287,** 2185–2195
2. Hewes, R. S., and Taghert, P. H. (2001) *Genome Res.* **11,** 1126–1142
3. Vanden Broeck, J. (2001) *Peptides* **22,** 241–254
4. De Loof, A., Baggerman, G., Breuer, M., Claeys, I., Cerstiaens, A., Clynen, E., Janssen, T., Schoofs, L., and Vanden Broeck, J. (2001) *Arch. Insect Biochem. Physiol.* **47,** 129–138
5. Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001) *Curr. Biol.* **11,** 213–221
6. Baggerman, G., Huybrechts, J., Clynen, E., Hens, K., Harthoorn, L., Van der, H. D., Poulos, C., De Loof, A., and Schoofs, L. (2002) *Peptides* 635–644
7. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) *J. Mol. Biol.* **300,** 1005–1016
8. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng.* **10,** 1–6
9. Schneider, L. E., and Taghert, P. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 1993–1997
10. Nichols, R. (1992) *J. Mol. Neurosci.* **3,** 213–218
11. Nichols, R., Schneuwly, S. A., and Dixon, J. E. (1988) *J. Biol. Chem.* **263,** 12167–12170
12. Wei, Z., Baggerman, G., Nachman, J., Goldsworthy, G., Verhaert, P., De Loof, A., and Schoofs, L. (2000) *J. Insect Physiol.* **46,** 1259–1265
13. Schaffer, M. H., Noyes, B. E., Slaughter, C. A., Thorne, G. C., and Gaskell, S. J. (1990) *Biochem. J.* **269,** 315–320
14. Bogerd, J., Kooiman, F. P., Pijnenburg, M. A., Hekking, L. H., Oudejans, R. C., and Van Der Horst, D. J. (1995) *J. Biol. Chem.* **270,** 23038–23043
15. Oudejans, R. C., Mes, T. H., Kooiman, F. P., and Van Der Horst, D. J. (1993) *Peptides* **14,** 877–881
16. Veenstra, J. A. (1989) *FEBS Lett.* **250,** 231–234
17. Veenstra, J. A. (1994) *Biochem. Biophys. Res. Commun.* **204,** 292–296
18. Lenz, C., Williamson, M., and Grimmelikhuijzen, C. J. (2000) *Biochem. Biophys. Res. Commun.* **273,** 1126–1131
19. Williamson, M., Lenz, C., Winther, A. M., Nassel, D. R., and Grimmelikhuijzen, C. J. (2001) *Biochem. Biophys. Res. Commun.* **282,** 124–130
20. Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and De Loof, A. (1991) *Regul. Pept.* **36,** 111–119
21. Huesmann, G. R., Cheung, C. C., Loi, P. K., Lee, T. D., Swiderek, K. M., and Tublitz, N. J. (1995) *FEBS Lett.* **371,** 311–314
22. Davies, S. A., Stewart, E. J., Huesmann, G. R., Skaer, N. J., Maddrell, S. H., Tublitz, N. J., and Dow, J. A. (1997) *Am. J. Physiol.* **273,** R823–R827
23. Schoofs, L., Holman, G. M., Paemen, L., Veelaert, D., Amelinckx, M., and De Loof, A. (1993) *Peptides* **14,** 409–421
24. Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and De Loof, A. (1991) *Gen. Comp Endocrinol.* **81,** 97–104
25. Torfs, P., Nieto, J., Cerstiaens, A., Boon, D., Baggerman, G., Poulos, C., Waelkens, E., Derua, R., Calderon, J., De Loof, A., and Schoofs, L. (2001) *Eur. J. Biochem.* **268,** 149–154
26. Schoofs, L., Veelaert, D., Vanden Broeck, J., and De Loof, A. (1997) *Peptides* **18,** 145–156
27. Terhzaz, S., O'Connell, F. C., Pollock, V. P., Kean, L., Davies, S. A., Veenstra, J. A., and Dow, J. A. (1999) *J. Exp. Biol.* **202,** 3667–3676
28. Perone, M. J., and Castro, M. G. (1997) *Histol. Histopathol.* **12,** 1179–1188
29. Uttenweiler-Joseph, S., Moniatte, M., Lageux, M., Van Dorsselaer, A., Hoffmann, J. A., and Bulet, P. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11342–11347
30. Nathoo, A. N., Moeller, R. A., Westlund, B. A., and Hart, A. C. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 14000–14005
31. 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) *J. Biol. Chem.* **275,** 23273–23280
32. Veelaert, D., Passier, P., Devreese, B., Vanden Broeck, J., Van Beeumen, J., Vullings, H. G., Diederen, J. H., Schoofs, L., and De Loof, A. (1997) *Endocrinology* **138,** 138–142
33. Nassel, D. R., Shiga, S., Mohrherr, C. J., and Rao, K. R. (1993) *J. Comp. Neurol.* **331,** 183–198
34. Zitnan, D., Kingan, T. G., Hermesman, J. L., and Adams, M. E. (1996) *Science* **271,** 88–91
35. Phlippen, M. K., Webster, S. G., Chung, J. S., and Dircksen, H. (2000) *J. Exp. Biol.* **203 Pt 3,** 521–536
36. Horodyski, F. M., Ewer, J., Riddiford, L. M., and Truman, J. W. (1993) *Eur. J. Biochem.* **215,** 221–228
37. Nambu, J. R., Murphy-Erdosh, C., Andrews, P. C., Feistner, G. J., and Scheller, R. H. (1988) *Neuron* **1,** 55–61
38. Kean, L., Pollock, V. P., Broderick, K. E., Dalies, S. A., Veenstra, J., and Dow, J. A. T. (2002) *Am. J. Physiol.* **282,** 1297–1307

Peptidomics of the Larval *Drosophila melanogaster*Central Nervous System

Geert Baggerman, Anja Cerstiaens, Arnold De Loof and Liliane Schoofs

J. Biol. Chem. 2002, 277:40368-40374.
doi: 10.1074/jbc.M206257200 originally published online August 8, 2002
