Neuropeptide precursor gene discovery in the Chagas disease vector *Rhodnius prolixus*

S. Ons*, M. Sterkel*, L. Diambra†, H. Urlaub‡ and R. Rivera-Pomar*§

**Laboratorio de Genética y Genómica Funcional, Centro Regional de Estudios Genómicos, Universidad Nacional de La Plata, Buenos Aires, Argentina; †Laboratorio de Biología de Sistemas, Centro Regional de Estudios Genómicos, Universidad Nacional de La Plata, Buenos Aires, Argentina; ‡Bioanalytical Mass Spectrometry Laboratory, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany; and §Departamento de Ciencias Básicas y Experimentales, Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Pergamino, Argentina*

Abstract

We show a straightforward workflow combining homology search in *Rhodnius prolixus* genome sequence with cloning by rapid amplification of cDNA ends and mass spectrometry. We have identified 32 genes and their transcripts that encode a number of neuropeptide precursors leading to 194 putative peptides. We validated by mass spectrometry 82 of those predicted neuropeptides in the brain of *R. prolixus* to achieve the first comprehensive genomic, transcriptomic and neuropeptidomic analysis of an insect disease vector. Comparisons of available insect neuropeptide sequences revealed that the *R. prolixus* genome contains most of the conserved neuropeptides in insects, many of them displaying specific features at the sequence level. Some gene families reported here are identified for the first time in the order Hemiptera, a highly biodiverse group of insects that includes many human, animal and plant disease agents.

Keywords: insect neuropeptides, peptidomic, transcriptomic, genomic, *Rhodnius prolixus*.

Introduction

Multicellular organisms, such as insects, require signalling molecules for cell-to-cell communication. Amongst these molecules, the neuropeptides play a central role in the control of development, reproduction, behaviour, feeding and many other physiological processes. Neuropeptides are produced in endocrine cells or neurones as large precursors. These precursors are cleaved and further modified to yield mature peptides that are secreted to the extracellular environment. Often, a single precursor encodes several biologically active molecules, which can be readily predicted from the precursor gene, as proteolytic processing occurs at specific mono and dibasic cleavage sites (Nassel, 2002).

New techniques have recently emerged to identify the genes encoding neuropeptides. These approaches include bioinformatic tools to predict neuropeptide precursor genes (NPGs) from genomes, and direct detection of the processed mature peptides using mass spectrometry (MS) (Baggerman *et al.*, 2002; Hummon *et al.*, 2006; Li *et al.*, 2008; Reumer *et al.*, 2008). Upon identification of a putative neuropeptide-encoding gene from a genomic sequence, the simplest way to validate the bioinformatic prediction is the sequencing of rapid amplification of cDNA ends (RACE) products, which provide information about gene expression. RACE-PCR allows the identification of peptide precursor mRNAs from incomplete or not assembled genomes. The amino acid sequence of the mature peptides obtained by means of proteomics, which also includes post-translational modifications, can be compared with the translated open reading frames (ORFs) of the mRNAs, revealing the origin of the mature peptides. As such, *de novo* gene prediction, transcriptomics and peptidomics are complementary methods and the prerequisites for further studies on the functional genomics of neuropeptides.

The insect neuropeptidome is a promising target for a novel generation of insecticides that offer improved selectivity and environmental compatibility (Scherkenbeck & Zdobinsky, 2009). The knowledge of the neuropeptide set and the understanding of their function, active conformations and interactions, provides the tools for the design of peptidomimetics, pseudopeptides or small molecules capable of disrupting the physiological processes regulated by the signalling molecules and their receptors. Thus, the structural and functional characterization of the neuropeptidome of insects of economical interest is the first requirement to develop strategies to replace or complement conventional neurotoxic insecticides.

The triatomine *Rhodnius prolixus* is one of the main vectors of Chagas disease, which is caused by the protozoan *Trypanosoma cruzi*. Chagas disease is an important human disease, with 10 million people affected in Central and South America, and another 120 million (one quarter of the population of the region) at risk (www.who.int/mediacentre/Factsheets/FS340/en/). Owing to the lack of vaccines and effective treatments for the chronic stage of the disease, the control of vector populations has become the preferred management strategy. However, the use of conventional insecticides in the endemic areas has resulted in the rise of insecticide resistant populations (Picollo *et al.*, 2005). The genome of *R. prolixus* is in an advanced sequencing stage. Although it is not yet fully assembled or annotated, the need to move forward to the identification of potential targets and/or novel control mechanisms such as the design of peptidomimetics, led us to investigate the neuropeptide set of this vector. It is noteworthy that the experimental identification of the neuropeptidome of disease vectors is just at the beginning (Neupert *et al.*, 2009; Ons *et al.*, 2009; Predel *et al.*, 2010). In the case of *R. prolixus*, the sequence of only two NPGs has been reported until now (Paluzzi *et al.*, 2008; Neupert *et al.*, 2010; 2010).

Here we combine bioinformatics, transcriptomics and proteomics to determine a large set of neuropeptide precursors in *R. prolixus*. Homology and peptidomic-based search in genomic trace archives of whole sequence genome, shotgun contigs and expressed sequence tags (ESTs) databases of *R. prolixus*, cDNA based cloning and MS, served to identify and validate the expression of new neuropeptide precursor genes in *R. prolixus* neuroendocrine system. A combined approach as we present here has not been reported for any disease vector, even those with fully sequenced genomes and, in addition, it contributes to a better understanding of the order Hemiptera, a highly biodiverse group that includes many species of insects harmful for human, animals and plants.

Results and discussion

Neuropeptide precursor genes: transcriptomic and proteomic validation of the bioinformatic prediction

In order to identify NPGs in the *R. prolixus* genome, we performed homology searches on *R. prolixus* whole genome trace archives, shotgun sequence contigs and EST databases. We performed TBLASTN searches using sequences derived from other arthropod species or neuropeptide sequences obtained by MS-driven *de novo* sequencing of *R. prolixus* (Ons *et al.*, 2009) as queries. Figure 1 shows the strategy of *in silico* search and validation process.

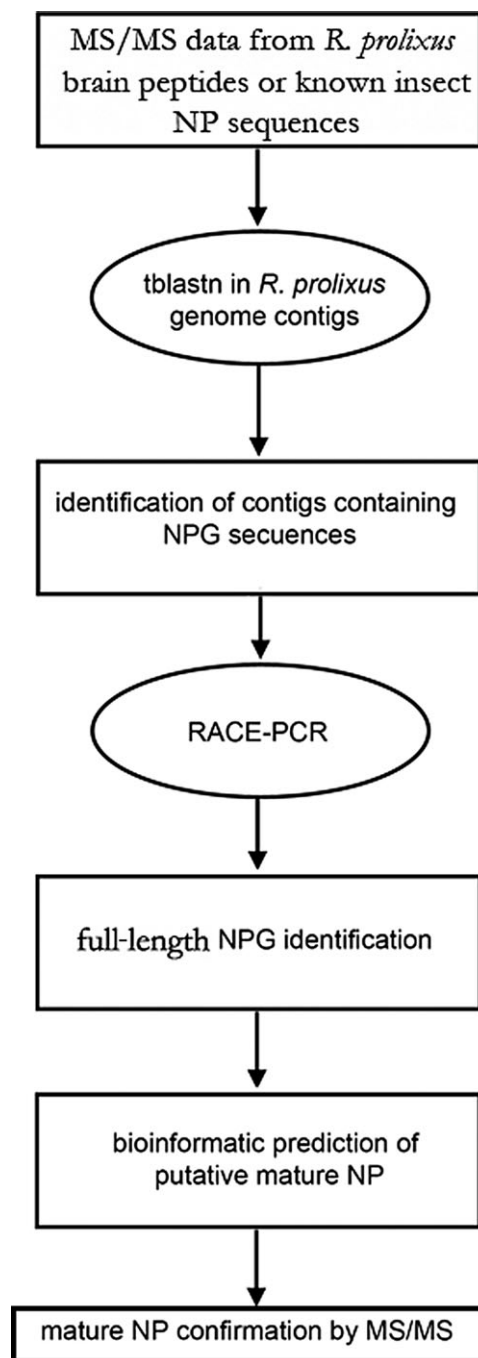This strategy yielded, in many cases, incomplete scaffolds or small pieces of the genome. In order to identify the

Figure 1. Experimental approach and flow search used to the discovery of the neuropeptide precursor genes (NPGs). MS/MS, tandem mass spectrometry; RACE, rapid amplification of cDNA ends.

complete transcripts from the unassembled genomic information and, from these, determine the gene structure, we performed RACE-PCR on cDNA from the nervous system of *R. prolixus*. Forward and reverse primer sets were designed from the partial sequences identified in genomic traces and contigs (Table S1). The RACE products were cloned and sequenced. This technique,

although laborious, brought more reliable transcript sequences than the sole bioinformatics predictions, and confirmed the expression of the genes in the nervous system. When the neuropeptide-encoding transcript was identified and sequenced, the encoded pre-propeptide was predicted by an ORF search, the number of putative peptides that could be produced by proteolysis determined by bioinformatics analysis, and, finally, validated by tandem MS analysis. We determined the complete cDNA sequence of most of the mRNAs derived from the NPGs, although a minor number of transcripts [eclosion hormone (EH), insect kinin, pyrokinin (PK) and FIRF-amide related peptide] could not be cloned or fully sequenced and, thus, they were only predicted by bioinformatics means. However, MS data confirmed the presence of most *in silico* predicted NPG-encoded mature peptides (Table 1). Of these, a few peptides, including two allatostatins, long neuropeptide F and six tachykinins, showed minor differences to the sequence obtained by *de novo* sequencing (Ons *et al.*, 2009). The genomic and transcriptomic analysis performed here confirmed the sequence presented here.

Based on the complete transcript sequence we assembled the available genomic sequence and determine the gene structure. Figure 2 summarizes the predicted gene structure of 22 NPGs, including the splicing variants determined for two of them (calcitonin-like diuretic hormone and ion transport peptide, ITP). Along with the full NPG transcripts, we identified the partial sequence of 10 additional NPGs, which, along with the two CAPA genes recently reported (Paluzzi *et al.*, 2008; Neupert *et al.*, 2010; Paluzzi & Orchard, 2010) adds up a total number of 34 NPGs in *R. prolixus*. The annotation of the fully assembled genomes of *Bombyx mori, Tribolium castaneum, Drosophila melanogaster, Anopheles gambiae* and *Apis mellifera* has determined 37, 47, 31, 32 and 36 NPGs, respectively (Hewes & Taghert, 2001; Riehle *et al.*, 2002; Hummon *et al.*, 2006; Li *et al.*, 2008; Roller *et al.*, 2008). Therefore, the 34 NPGs of *R. prolixus* are in the expected range, indicating that our approach has been reliable for the discovery of most of the NPGs. The comparison of the occurrence of the different NPGs shows wide conservation, with only one transcript being *R. prolixus*-specific. The data are summarized in Table 1.

Based on the information described above we used bioinformatics tools (see Experimental procedures and http://neuroproteomics.scs.illinois.edu/neuropred.html) and manual analysis and correction to determine the possible number of peptides that could derive after the processing of the encoded pre-propeptides. The peptides were predicted in sequences located between putative cleavage sites. Amongst the predicted peptides, several were identified either with or without post translational modifications, namely oxidation of Met, N-terminal acety-

lation of Ala and N-terminal pyro-glutamic of Gln. In some cases multiple truncated peptides were detected from a single precursor. As fragments were found in different chromatographic fractions, this cannot be a consequence of fragmentation during the ionization process. The same truncated forms were observed in different experimental replicas and the short peptide forms were observed in several specific neuropeptides, but not in others. We cannot rule out the possibility that this was a consequence of the extraction process, but it seems to be a common observation in other peptidomic studies performed with different extraction methods (Hummon *et al.*, 2006; Predel *et al.*, 2008). The same consideration applies for oxidized and non-oxidized forms of some peptides. All these points suggest that the truncated and the oxidized forms normally occur in the brain. With these considerations, we predict 194 *bona fide* mature neuropeptides, and validated 82 of them in the brain of *R. prolixus* by means of MS and tandem MS (MS/MS) with a cut-off of 4 kDa for precise analysis, which excludes several of the predicted ones. Table 2 summarizes the data of the encoded ORFs, the putative mature products and the peptides validated.

Comparative analysis of neuropeptides

Novel variants of the extended RF-amide peptides. In insects, a variety of neuropeptides share the C-terminal RF-amide motif. These include extended FMRF-amides, sulphakinins (SK), myosuppressins (MSP), short (sNPF) and long (LNF) neuropeptide F. We found all these genes in *R. prolixus* nervous tissue, which shows specific sequence attributes, unique amongst insects. Interestingly, in the highly conserved C-terminal region of *R. prolixus* MSP the active site shows a FMRF-NH_2 (-NH_2=C terminal amidation), instead of the FLRF-NH_2-C terminal domain, which is characteristic of the MSPs described so far (Nassel, 2002; Orchard & Lange, 2006). The SK precursor sequence consists of eight predicted peptides. One of the *R. prolixus* SK peptides, GGSDEKFDDYGYMRF-NH_2, is distinctive amongst known insect SKs, as the C-terminal sequence GYMRF-NH_2 is unique amongst known SKs because of its GHMRF-NH_2 C-terminus (Nassel, 2002; Schoofs & Nachman, 2006).

The *R. prolixus* RF-amide precursor shows acetylation in one of the mature FIRF-NH_2 peptides (Acetyl-AKDNFIRF-NH_2) a unique feature amongst the members of this family described so far (Nassel, 2002; Orchard & Lange, 2006).

We found the sequence of a gene encoding sNPF. This sequence contains three putative mature peptides, flanked by pairs of basic residues. From these, the only one related with the sNPF family is NNRSPQLRLRF-NH_2, although NDPTFLQGDHLMDNSMIDTL was also

Table 1. Neuropeptide precursors identified in *Rhodnius prolixus* genomic information

Name	GenBank accession no.	mRNA*	MS†	Closest BLASTP match in SwissPROT	Tribolium castaneum	Drosophila melanogaster	Anopheles gambiae	Apis mellifera	Bombyx mori
Adipokinetic hormone	ID: 1273228	+	−	Hypertrehalosaemic peptide *Blaberus discoidalis* ($E = 3e{-}05$).	+	+	+	+	+
Allatostatin A	GQ856315	+	8/16	Allatostatin *Diploptera punctata* ($E = 4\ e{-}26$).	−	+	+	+	+
Allatostatin B	ID: 1273333	+	9/12	Prothoracicostatic hormone *Bl.mori* ($E = 5\ e{-}29$).	+	+	+	−	+
Allatotropin	GQ162783	+	3/6	Allatotropin precursor *Manduca sexta* ($E = 4\ e{-}09$).	+	−	+	+	+
Apis prohormone 1	ID: 1273236	+	−	Prohormone 1 *Ap. mellifera* ($E = 2e{-}21$).	−	+	+	+	−
Calcitonin-like diuretic hormone A	GQ856316	+	2/4	CT-like DH *D. melanogaster* ($E = 4\ e{-}12$).	+	−	−	+	−
Calcitonin-like diuretic hormone B	GQ856317	+	2/3	CT-like DH *D. melanogaster* ($E = 4\ e{-}19$).	+	−	+	+	+
Calcitonin-like diuretic hormone C	ID: 1273251	+	−	No significant match	−	−	+	−	−
CCH-amide like peptide	ID: 1273254	+	(1/3)	No significant match‡	+	+	+	+	+
Corazonin	ID: 1273265	+	1/3	Corazonin precursor *Bo. mori* ($E = 6\ e{-}10$).	−	+	+	+	+
Crustacean cardioactive peptide	ID: 1273341	+	1/7	Cardioactive peptide *Periplaneta americana* ($E = 4e{-}28$).	+	+	+	+	+
Eclosion hormone	ID: 1273270	−	(1/2)	Eclosion hormone *Manduca sexta* ($E = 3e{-}17$).	+	+	+	+	+
FMRF-amide related peptides	ID: 1273272	−	4/14	FMRFa- related peptides. *Drosophila virilis* ($E = 2\ e{-}12$).	+	+	+	+	+
Insect kinin	ID: 1273279	+	3/24	No significant match§	−	+	+	−	+
Ion transport peptide A	ID: 1273339	+	−	Ion transport peptide *Schistocerca gregaria* ($E = 1e{-}35$).	+	+	−	+	−
Ion transport peptide B	ID: 1273340	+	−	Ion transport peptide *S. gregaria* ($E = 4e{-}42$).	+	+	+	−	+
Long neuropeptide F	ID: 1273282	+	1/4	Neuropeptide Y *Aplysia californica* ($E = 4e{-}08$).	−	+	+	+	+
Myosuppressin	GQ344501	+	3/4	Myosuppressin *Ap. mellifera* ($E = 2\ e{-}09$).	+	+	+	+	+
Neuroparsin A	ID: 1273288	+	−	Neuroparsin A *Locusta migratoria* ($E = 2e{-}08$).	+	−	+	−	+
Neuropeptide-like precursor 1	ID: 1273296	+	9/20	Neuropeptide like precursor 1 *D. melanogaster* ($E = 0.058$)	+	+	+	+	+
Orcokinins	ID: 1273300	+	6/10	Orcokinin precursor *Procambarus clarkii* ($E = 4\ e{-}14$)	−	−	+	+	+
Pyrokinin	ID: 1273305	−	2/5	PBAN neuropeptides *Aedes aegypti* ($E = 0.006$)	+	+	+	+	+
Short neuropeptide F	ID: 1273338	+	2/3	Short Neuropeptide F *Ae. aegypti* ($E = 6\ e{-}08$)	+	+	+	+	+
SIF-amide	GQ253922	+	4/5	SIFa *Procambarus clarkii* ($E = 4\ e{-}12$)	+	+	+	+	+
Sulphakinins	GQ162784	+	5/8	Sulphakinins. *Calliphora vomitoria* ($E = 1\ e{-}07$)	+	+	+	+	+
Tachykinins	GQ162785	+	17/21	Tachykinins precursor *Ap. mellifera* ($E = 4\ e{-}11$)	+	+	+	+	+

*mRNA sequence of the precursor confirmed either by rapid amplification of cDNA ends PCR, cloning andsequencing, or in EST databases.

†Number of peptides identified by MS/total number of peptides predicted in the precursor.

Parentheses in the evidence for peptides are the ones confirmed only by matching mass.

‡Slight similarity with leukokinin from *B. mori* ($E = 1.6$ in nonredundant protein database).

§Slight similarity with CCH-amide from *B. mori* ($E = 0.014$ in nonredundant protein database).

CT, calcitonin; DH, diuretic hormone; *E*, expected value; EST, expressed sequence tag; MS, mass spectrometry; PBAN, pheromone biosynthesis activating neuropeptide.

Figure 2. Structure of the *Rhodnius prolixus* neuropeptide precursor genes. The numbers in each exon (black boxes) indicate its size in nucleotides. Contig number is indicated for each exon or group of exons. Lines represent introns. See text and tables for abbreviations.

detected in the brain (Table 2). The sNPF precursor is much shorter than the rest of the sNPF precursors described so far in insects.

Diuretic and antidiuretic peptides. Diuresis in insects is under the neurohormonal control of neuropeptides, and thus, diuretic hormones play an important role in the transmission of *T. cruzi* during post-prandial diuresis. We detected one antidiuretic hormone, ITP, and two diuretic hormones, CT-like DH and insect kinin. Alternative splicing has been reported in diuretic and antidiuretic hormones; ITP and CRF-like DH show regulated expression of different neuropeptides encoded by the same gene, in specific patterns (Dai *et al.*, 2007; Li *et al.*, 2008; Roller *et al.*, 2008). We determined two splicing variants for ITP in *R. prolixus*. These data provide evidence of the complex regulatory events of gene expression for the CT-like-DH

gene. *R. prolixus* appears to have at least three CT-like-DH mRNAs, produced by alternative splicing. The CT-like-DH gene consists of eight exons (Fig. 2). Isoforms A and B shares exons 1, 2, 4 and 8, whereas isoforms A and C shares exons 4, 5 and 6. Exons 3 and 7 only occur in isoform C. Exon 2 encodes the signal peptide, and exon 8 encodes for a mature peptide with diuretic activity (GLDLGLSRGFSGSQAAKHLMGLAAANYAGGP-NH$_2$; Brugge *et al.*, 2008; Te Brugge *et al.*, 2009). Isoforms A and B encode for secreted neuropeptides, from which at least one mature peptide displays diuretic activity, whereas isoform C encodes for a putative nonsecreted protein, in which the diuretic peptide is absent. The function of this isoform remains to be investigated. Interestingly, we have observed a tissue-specific expression pattern of the different splicing variants. Isoforms A and B are transcribed in nervous tissues, whereas isoform C was

Table 2. Sequences of neuropeptide precursors and mature peptides identified in *Rhodnius prolixus*

Consensus cleavage sites are underlined, *de novo* sequenced peptides are in bold, mass match identified peptides are double underlined, signal peptides, as predicted by SIGNALP, are in italics. Molecular mass of the protonated ion (M^+H^+) for each peptide is indicated.

Adipokinetic hormone (RhoprAKH gene)
MATNLFITSV LVLLTFHYTL AQLTFSTDWG KRSVRHNAPD CTPNPDTVIF LYKYLQNEFY KMIECGKTGG L

Encoded peptides:
QLTFSTDW-NH₂ (996.48)
pyro-QLTFSTDW-NH₂ (979.48)
SVRHNAPDCTPNPDTVIFLYKYLQNEFYKMIECGKTGGL (4462.17)

Allatostatin-A (RhoprAST-A gene)
MMLPFIVLLV VDVFALGAQA INDREDDFNK KLTELGLGKR AAYSYVSEYK RLPVYNFGLG KRAHNEGRLY SFGLGKRDYD SGEEMEYLDD ELAIRDELAK
RAAKMYSFGL GKRLPSIKYP EGKMYSFGLG KRVPFADQAY FLDDNDSSEE SKRSNPNGHR FSFGLGKRDE QEMNEKRKGE RSMQYSFGLG KRTQLDPANN
LHN

Encoded peptides:
INDREDDFN (1137.48)
LTELGL-NH₂ (644.40)
AAYSYVSEY (1052.46)
(Acetyl-A)AYSYVSEY (1094.47)
LPVYNFGL-NH₂ (921.51)
AHNEGRLYSFGL-NH₂ (1362.69)
(Acetyl-A)HNEGRLYSFGL-NH₂ (1404.70)
DYDSGEEMEYLDDELAIRDELA (2591.10)
AAKMYSFGL-NH₂ (986.52)
(Acetyl-A)AKMYSFGL-NH₂ (1028.52)
LPSIKYPEGKMYSFGL-NH₂ (1828.95)
VPFADQAYFLDDNDSSEES (2148.89)
SNPNGHRFSFGL-NH₂ (1331.66)
DEQEMNE (894.31)
GERSMQYSFGL-NH₂ (1273.60)
TQLDPANNLHN (1236.60)

Allatostatin-B (RhoprAST-B gene)
MSWCYKILLA TTLTAIVQGQ NPGTVTPMDN VIAEDEYLIP SNPALIDDKR SWKDLQSSGW GKRGWKDMQT VGWGKRAWTD LPSSGWGKKR AWSDLQSSGW
GKRGWKDMQS SGWGKRAWSD LQSSGWGKRA WSDLQSSGWG KRDWKDMQSS GWGKRAWSDL QSSGWGKRSK GDDQDEDIEE NLAEEDKRAW NSLHGGWGKR
TADWGSFRGS WGKREPAWQN LKGLWGKRSP VNLLQDNFQT GYVEPILN

Encoded peptides:
QNPGTVTPMDNVIAEDEYLIPSNPALIDD (3141.49)
pyro-QNPGTVTPMDNVIAEDEYLIPSNPALIDD (3124.49)
SWKDLQSSGW-NH₂ (1192.57)
GWKDMQTVGW-NH₂ (1206.57)
AWTDLPSSGW-NH₂ (1118.53)
DWKDMQSSGW-NH₂ (1238.53)
AWSDLQSSGW-NH₂ (1135.52)
SKGDDQDEDIEENLAEED (2050.82)
AWNSLHGGW-NH₂ (1026.49)
TADWGSFRGSW-NH₂ (1268.58)
EPAWQNLKGLWG-NH₂ (1397.73)
SPVNLLQDNFQTGYVEPILN (2261.14)

Allatotropin (RhoprAT gene)
*MMRWSSLLVL VALASIINCI KA*GSPSSALY SSAARASGRT RTIRGFKNVQ LSTARGFGKR TYPDSQLQPD LIPADWMAEE LSSNPELARF IIRRFIDVDQ
DGLVSPVELL RNTVCQEPN

Encoded peptides:
GSPSSALYSSAARASGRTRTIRGFKNVQLSTARGF-NH₂ (3656.95)
GFKNVQLSTARGF-NH₂ (1423.72)
FKNVQLSTARGF-NH₂ (1366.75)
NVQLSTARGF-NH₂ (1091.57)
TYPDSQLQPDLIPADWMAEELSSNPELARFII (3659.79)
FIDVDQDGLVSPVELLRNTVCQEPN (2800.38)

***Apis* Prohormone 1 (RhoprPH1 gene)**
MKVFAKMSSS QICAGVLFCL SMLIMLGMSQ PTPDKEKLLN ELSQELVEDD GSIDRAVIDY LYAKQLFNRL RAQAGAAEIQ QGKRSYWKQC AFNAVSCFGK

Encoded peptides:
QPTPDKEKLLNELSQELVEDDGSIDRAVIDYLYAKQLFNRLRAQAGAAEIQQ-NH₂ (5899.06)
pyro-QPTPDKEKLLNELSQELVEDDGSIDRAVIDYLYAKQLFNRLRAQAGAAEIQQ-NH₂ (5882.06)
SYWKQCAFNAVSCF-NH₂ (1652.73)

Table 2. *Continued*

Calcitonin-like diuretic hormone isoform A (RhoprCTDH gene)

*MVTNIAVVGV SLMLGTLIVL SAA*SENIPYI GHRASYFGDM DNEPDSEVML EILAKLGRTI MRANDLEKPM IYSREASNPW TAVNKLRPSN LPYNIELAEN PDSIYS<u>KRGL</u> DLGLSRGFSG SRAAKHLMGL AAANYAGGPG RR<u>RQA</u>

Encoded peptides:

SENIPYIGHRASYFGDMDNEPDSEVMLEILAKLGRTIMRANDLEKPMIYSREASNPWTAVNKLRPSNLPYNIELAENPDSIYS (9469.63)

GLDLGLSRGFSGSQAAKHLMGLAAANYAGGP-NH₂ (2986.53)

GLDLGLSRGFSGSQAA (1535.79)

Calcitonin-like diuretic hormone isoform B (RhoprCTDH gene)

*MVTNIAVVGV SLMLGTLIVL SAA*SENIPYI GHRASYFGDM DNEPDSEVML EILAKLGRTI MRANDLENS<u>K</u> <u>R</u>GLDLGLSRG FSGSQAAKHL MGLAAANYAG GPGRR<u>RQA</u>

Encoded peptides:

SENIPYIGHRASYFGDMDNEPDSEVMLEILAKLGRTIMRANDLENS (5211.48)

GLDLGLSRGFSGSQAAKHLMGLAAANYAGGP-NH₂ (2986.53)

GLDLGLSRGFSGSQAA (1535.79)

Calcitonin-like diuretic hormone isoform C (RhoprCTDH gene)

MDNEPDSEVM LEILAKLGRT IMRANDLEKP MIYSREASNP WTAVNKLRPS NLPYNVELAE NPDSI

No cleavage sites predicted in the protein.

CCH-amide (RhoprCCH gene)

*MICSRKMIVT LLLVSLLLTV HG*AAFKGARD GDASF<u>RKK</u>PL <u>RR</u>GGCSAFGH SCFGGHG<u>KRS</u> DDYMAQIQSR QLQRLPPADI VRQW

Encoded peptides:

<u>AAFKGARDGDASF</u> (1312.63)

<u>GG</u>CSAFGHSCFGGH-NH₂ (1322.51)

SDDYMAQIQSRQLQRLPPADIVRQW (3014.53)

Corazonin (RhoprCZ gene)

*MNFRSSCLLI FIIYSIVHVF G*QTFQYSRGW TNG<u>KR</u>AGIPS KEVTACQLQR IKSLLEGKTI PQLYWPCEWS PFMEAALSRQ MKTSELTSLP VVAPLTPEIE EK

Encoded peptides:

QTFQYSRGWTN-NH₂ (1386.65)

pyro-QTFQYSRGWTN-NH₂ (1369.64)

GIPSKEVTACQLQRIKSLLEGKTIPQLYWPCEWSPFMEAALSRQMKTSELTSLPVVA-PLTPEIEEK (7425.86)

Crustacean Cardioactive Peptide (RhoprCCAP gene)

*MQLLVPCFLL FTALVFA*VLT DDVFLQ<u>KR</u>VY FPGEIAEPID PKM<u>KK</u>PFCNAF TGCG<u>KKR</u>SDE SMATLVDLNS EPAVEELSRQ ILSEAKLWEA IQEARMELLN RKQQQSDRIP LQPLPLTTI<u>R</u> <u>KR</u>SHYLYT

Encoded peptides:

VLTDDVFLQ (1049.55)

VYFPGEIAEPIDPKM (1705.85)

PFCNAFTGC-NH₂ (956.39)

SDESMATLVDLNSEPAVEELSRQILSEAKLWEAIQEARMELLN (4858.41)

QQQSDRIPLQPLPLTTI (1948.09)

pyro-QQQSDRIPLQPLPLTTI (1931.86)

SHYLYT (783.37)

Eclosion hormone (RhoprEH gene)

*MKKLLLVILL TSFLAEISG*R QIGVCIRNCA QC<u>KK</u>MFGVYF EGQMCADTCL KYKGKLIPDC EDIASIGPFL NKL

Encoded peptides:

<u>R</u>QIGVCIRNCAQC (1463.70)

MFGVYFEGQMCADTCLKYKGKLIPDCEDIASIGPFLNKL (4358.11)

FMRF-amide Related Peptides (RhoprFMRFa gene)

*MMFGGMLVGA WLLIHGQCCV NA*GTDSRIRS PLVDPLI<u>RRS</u> PLEKNFMRFG <u>RS</u>SPALQQFP TAYNSNYLDL ENS<u>KRS</u>G<u>R</u>FD RARDNFMRFG <u>R</u>DNEKIALSN RAKDNFIRFG <u>R</u>SKDNFMRFG <u>R</u>IKDNFIRFG <u>R</u>GNDNFMRFG <u>R</u>SGDEIEEIL PKD<u>RR</u>DKALN RLG<u>R</u>QRLSDK SDNFIRFG<u>RL</u> VANFLLKIYF Y

Encoded peptides:

GTDSRIRSPLVDPLI (1638.92)

SPLEKNFMRF-NH₂ (1267.66)

SSPALQQFPTAYNSNYLDLENS (2459.14)

FDRARDNFMRF-NH₂ (1473.72)

DNEKIALSNRAKDNFIRF-NH₂ (2150.15)

AKDNFIRF-NH₂ (1009.55)

(Acetyl-A)KDNFIRF-NH₂ (1051.60)

SKDNFMRF-NH₂ (1043.51)

IKDNFIRF-NH₂ (1051.60)

GNDNFMRF-NH₂ (999.45)

SGDEIEEILPKD (1344.65)

DKALNRL-NH₂ (828.50)

QRLSDKSDNFIRF-NH₂ (1624.85)

LVANFLLKIYFY (1503.86)

Table 2. *Continued*

Insect kinin (RhoprKN gene)

*MILLWMVWTI AVICKTSQ*GN DIISTSAEGH NLTTAPSPLP TAKDNSKGIR D<u>KR</u>GTGHLLE QLLKENELAA EDLEDEEDLV NDKI<u>KR</u>TNNR GNFAGNPRMR
FSSWAG<u>KR</u>AK FSSWGG<u>KR</u>VD DELISGTDGP IEYEIPED<u>KR</u> ANKFSSWAG<u>K</u> <u>R</u>TDEEGVNWM GNSPADLDSF IQQLEQ<u>KR</u>AK FSSWAG<u>KR</u>DE DRQKFSHWAG
<u>KK</u>FDDSLNMN DVLLEEE<u>KR</u>G AKFSSWAG<u>KR</u> AKFNSWGG<u>KR</u> FSNEFMNDNN DIEKNIVEE<u>K</u> <u>R</u>LSINPW<u>KK</u>I DDNG<u>KR</u>AKFS SWGG<u>KR</u>ADDD WL<u>KK</u>ARFNSW
GG<u>KR</u>NSFNAN ITNSVDDLFL DHEDALI<u>KRS</u> AAAYTPLSW<u>K</u> <u>R</u>KPIFSSWGG <u>KR</u>TARSTQPQ <u>RR</u>LIFPSNLF RDHSTWGALL RPI<u>RR</u>GPDFY AWGG<u>KR</u>ST

Encoded peptides:
NDIISTSAEGHNLTTAPSPLPTAKDNSKGIRD (3320.67)
GTGHLLEQLLKENELAAEDLEDEEDLVNDKI (3492.72)
TNNRGNFAGNPRMRFSSWA-NH$_2$ (2182.05)
AKFSSWG-NH$_2$ (781.40)
VDDELISGTDGPIEYEIPED (2205.99)
ANKFSSWA-NH$_2$ (909.46)
TDEEGVNWMGNSPADLDSFIQQLEQ (2823.24)
AKFSSWA-NH$_2$ (795.41)
DEDRQKFSHWA-NH$_2$ (1417.66)
FDDSLNMNDVLLEEE (1782.77)
GAKFSSWA-NH$_2$ (852.44)
AKFNSWG-NH$_2$ (808.41)
FSNEFMNDNNDIEKNIVEE (2300.99)
LSINPW (729.39)
IDDN-NH$_2$ (475.21)
AKFSSWG-NH$_2$ (781.40)
ADDDWL (734.30)
ARFNSWG-NH$_2$ (836.42)
NSFNANITNSVDDLFLDHEDALI (2577.21)
SAAAYTPLSW (1066.52)
PIFSSWG-NH2 (792.40)
TARSTQPQ (888.45)
LIFPSNLFRDHSTWGALLRPI **(**2453.34)
GPDFYAWG-NH2 (911.40)

Ion transport peptide isoform A (RhoprITPA gene)

MHQERRALAG LVVASTLLSW AVAGPSSRLV LSHPLN<u>KR</u>SF FDLQCKGVYD KSIFARLDRI CEDCYNLFRE PQLHSLCRSD CFASKYFAGC LEALLLREEE
NKFFQMVEFL G

Encoded peptides:
GPSSRLVLSHPLN (1376.76)
SFFDLQCKGVYDKSIFARLDRICEDCYNLFREPQLHSLCRSDCFASKYFAGCLEALLLREEENKFFQMVEFL-NH$_2$ (8586.12)

Ion transport peptide isoform B (RhoprITPB gene)

MHQERRALAG LVVASTLLSW AVAGPSSRLV LSHPLN<u>KR</u>SF FDLQCKGVYD KSIFARLDRI CEDCYNLFRE PQLHSLC<u>RKN</u> CFTTDYFKGC LEVLLLQDEM
ENIQTWIKQL HGAEPEV

Encoded peptides:
GPSSRLVLSHPLN (1376.76)
SFFDLQCKGVYDKSIFARLDRICEDCYNLFREPQLHSLC (4672.22)
NCFTTDYFKGCLEVLLLQDEMENIQTWIKQLHGAEPEV (4455.13)

Long neuropeptide F (RhoprLNF gene)

*MNCWLLWLWT GLMACNAAMA M*AQPIPADAM ARPARPKSFA SPDDLRTYLD QLGQYYAVAG RPRFG<u>KR</u>AGG INPRLHLAVD GVNYRYPLAD ASDLYDLLFQ
QQSTE

Encoded peptides:
QPIPADAMARPARPKSFASPDDLRTYLDQLGQYYAVAGRPRF-NH$_2$ **(**4705.42)
pyro-QPIPADAMARPARPKSFASPDDLRTYLDQLGQYYAVAGRPRF-NH$_2$ (4688.42)
AVAGRPRF-NH$_2$ (872.53)
AGGINPRLHLAVDGVNYRYPLADASDLYDLLFQQQSTE (4220.10)

Myosuppressins (RhoprMS gene)

*MILAWMCVTL LAGVL*AAPGP DCSPSALQVQ SSRVRNMCAL YQISSALQAY LDEQNNYQTA LRDTNIPYNI PE<u>KR</u>QDIDHV FMRFGRRR

Encoded peptides:
APGPDCSPSALQVQSSRVRNMCALYQISSALQAYLDEQNNYQTALRDTNIPYNIPE **(**6213.97)
QDIDHVFMRF-NH$_2$ (1306.64)
pyro-QDIDHVFMRF-NH$_2$ (1289.60)
pyro-QDIDHVF(M-OH)RF-NH$_2$ (1305.60)

Neuroparsin A (RhoprNPA gene)

*MSSQSSKTAT TALAVLTIFC MVALVSG*VFY GCVPCIGDEC NLNPGNCPYG IVRDPCGRLV CAAGPGERCG GRDFHLGKCG EGLSCKCGKC RGCSIKQIMN
GRIDCDTTNP MCQ

Encoded peptides:
VFYGCVPCIGDECNLNPGNCPYGIVRDPCGRLVCAAGPGERCGGRDFHLGKCGEGLSCKCGKCRGCSIKQIMNGRIDCDTTNPMCQ

Table 2. *Continued*

Neuropeptide like precursor 1 (RhoprNP1 gene)

MIATSLAPLI LTLLLLSKAS GENNDNTKSS SKHALVTDTG EEHNLEKRHV SSLLGNRASG PYQTGKRSSP SKSSLDELAE RLEEAAQEDK RYLGALARSG DLRVVARDRQ EKREDLDSLI DELASTEEMR RMQFDALRDD LFEEEPDKRG VVSLARAGYL KPTTHDFLED DEESSFYPAE DEDKRGGIAS LARNGYYQKR TVDAELEQLM SEVYGIGEKR SVASLARSYN LPNAVKGGYE NDDEKRNIPS LLRDRTSPLG EGKRHIGSFV ANHGIPFVNN KEGGKRSVGS LARNRDPFYA VKFGKRDAPD EEGEEMSKRY VATLLRQGRL PIGIDSPDHG EMSSMKEDND DHDVDKDEMS QVMQDEASKL SIRKKKSVPT VEGPTRIKRE AAADEFGGVR DDDSLAQFAD DAADSPINKR YFGVRGGGKM PGGRLPKVGG RSRNRHENSG RRRH

Encoded peptides:

NNDNTKSSSKHALVTDTGEEHNLE (2640.21)
HVSSLLGNRASGPYQT-NH₂ (1685.87)
SSPSKSSLDELAERLEEAAQED (2391.12)
YLGALARSGDLRVVARDRQE (2245.22)
EDLDSLIDELASTEEM (1809.80)
MQFDALRDDLFEEEPD (1969.85)
GVVSLARAGYLKPTTHDFLEDDEESSFYPAEDED (3802.72)
GGIASLARNGYYQ (1369.69)
TVDAELEQLMSEVYGIGE **(**1982.93)
SVASLARSYNLPNAVKGGYENDDE (2569.22)
NIPSLLRDRTSPLGE-NH₂ (1666.92)
HIGSFVANHGIPFVNNKEG-NH₂ (2036.05)
SVGSLARNRDFPYAVKF-NH₂ (1926.03)
DAPDEEGEEMS (1208.43)
YVATLLRQ-NH₂ (962.58)
LPIGIDSPDHGEMSSMKEDNDDHDVDKDEMSQVMQDEASKLSI (4788.09)
SVPTVEGPTRI (1155.63)
EAAADEFGGVRDDDSLAQFADDAADSPIN (2982.29)
YFGVRGGGKMPGGRLPKVG-NH₂ (1932.07)
SRNRHENS-NH₂ (998.49)

Orcokinin (RhoprOK gene)

MINMLSLTIL AMAVAVTSAF PRGELGVEEG NLYPGLYRDQ TMEDKEGRNL DTLGSGNLLR DLEAVLRAHP NLFYGRPARN HDTLDSLSGI TFGSQKRFDP LSSAYAADKR NFDEIDRSGF NSFIKKKKNF DEIDRSGFDGF VKRNFDEIDR VGFGSFIKK

Encoded peptides:

FPRGELGVEEGNLYPGLYRDQTMEDKE-NH₂ (3141.49)
NLDTLGSGNLLRDLEAVLAHPNLFY-NH₂ (2754.46)
PARNHDTLDSLSGITFGSQ (2015.98)
FDPLSSAYAAD (1156.52)
NFDEIDRSGFNSFI (1660.77)
NFDEIDRSGFN (1313.58)
NFDEIDRSGFDGFV (1617.73)
NFDEIDRSGFDG (1371.59)
NFDEIDRVGFGSFI (1615.74)
NFDEIDRVGF (1211.57)

Pyrokinin (RhoprPK gene)

MVSVSLVGLL LVALQLITNG CTQEGGRNTV NFSPRLGRDE EVVFTETSRS PPFAPRLGRI VFRPRFGRLT LAAQH

Encoded peptides:

EGGRNTVNFSPRL-NH₂ (1445.76)
DEEVVFTETSRSPPFAPRL-NH₂ (2176.10)
SPPFAPRL-NH₂ (883.51)
IVFRPRF-NH₂ (933.58)
LTLAAQH (753.43)

Short neuropeptide F (Rhopr sNPF gene)

MKIALSALCC LIAVALMFTP ETTSAPAIQD YDSMRDLYEL LLQREALPDS WAHKVVRKNN RSPQLRLRFG RRNDPTFLQG DHLMDNSMID TL

Encoded peptides:

APAIQDYDSMRDLYELLLQREALPDSWAHKVV (3742.89)
NNRSPQLRLRF-NH₂ (1399.79)
NDPTFLQGDHLMDNSMIDTL (2277.02)

SIF-amide (RhoprSIFa gene)

MSRTLFVCCF TLVVALIFLD AAMATYKKPP FNGSIFGKRA GPSSDYETAG KALSTMCEIA AEACSAWFPV QDNN

Encoded peptides:

TYKKPPFNGSIF-NH₂ (1397.74)
YKKPPFNGSIF-NH₂ (1296.70)
KKPPFNGSIF-NH₂ (1133.66)
KPPFNGSIF-NH₂ (1005.56)
AGPSSDYETAGKALSTMCEIAAEACSAWFPVQDNN **(**3634.58)

Sulphakinins (RhoprSK gene)

MGSSFLITLL LAIGVYMFIE NSHFMCLAEP AERRSLIRIR PEPALFAAED DPLDIVDKRQ FNEYGHMRFG KRGGSDEKFD DYGYMRFGRS RPLANSLPN

Table 2. *Continued*

Encoded peptides:
SHFNCLAEPAE (1217.52)
SLIRIRPEPALFAAEDDPLDIVD (2565.36)
QFNEYGHMRF-NH₂ (1327.60)
pyro-QFNEYGHMRF-NH₂ (1310.60)
pyro-QFNEYGH(M-OH)RF-NH₂ (1326.60)
GGSDEKFDDYGYMRF-NH₂ (1785.75)
GGSDEKFDDYGY(M-OH)RF-NH₂ (1801.76)
SRPLANSLPN (1068.58)

Tachykinin (RhoprTK gene)
MPVGSLLVMS CVLAACLAQE RRAMGFVGMR GKKDTPDMEE YKRAPSTMGF QGVRGKKDDL IGEPDDTFLE EFKRAPAAMG FQGMRGKKTP AMGFMGMRGK KDSDYGWWEE DKRAPASGFF GMRGKKAPAS GFFGMRGKKG PSSSAFFGMR GKKGPSGFMG VRGKKDSPDD LNHLLQLLRE SALKQEMEEM LEDGRGLKRF AGLSDSFEDY PAELL

Encoded peptides:
QERRAMGFVGMR-NH₂ (1436.73)
pyro-QERRAMGFVGMR-NH2 (1419.72)
DTPDMEEY (999.36)
STMGFQGVR-NH₂ (981.49)
APSTMGFQGVR-NH₂ (1149.53)
(Acetyl-A)PSTMGFQGVR-NH₂ (1191.59)
DDLIGEPDDTFLEEF (1754.76)
APAAMGFQGMR-NH₂ (1135.51)
(Acetyl-A)PAAMGFQGMR-NH₂ **(**1177.56)
TPAMGFMGMR-NH₂ (1097.48)
DSDYGWWEED (1301.45)
APASGFFGMR-NH₂ (1039.52)
ASGFFGMR-NH₂ (871.42)
ASGFFG(M-OH)R-NH₂ (887.42)
SGFFGMR-NH₂ (800.38)
(Acetyl-A)PASGFFGMR-NH₂ **(**1081.52)
GPSSSAFFGMR-NH₂ (1142.53)
GPSGFMGVR-NH₂ (906.46)
FMGVR-NH₂ (608.33)
DSPDDLNHLLQLLRESALKQEMEEMLEDGRGL (3694.80)
FAGLSDSFEDYPAELL (1773.82)

the only variant observed in ESTs derived from testis. One CAPA transcript, although related to the control of diuresis, has been also found in *R. prolixus* testes (Paluzzi *et al.*, 2008). Our finding suggests a novel function for a protein encoded by the CT-like-DH gene, probably involved in reproduction.

Insect kinin neuropeptides share a common C-terminal pentapeptide sequence $FX_1X_2WG-NH_2$ (X_1 = F, H, N, S, Y; X_2 = A, P, S; Nachman *et al.*, 2009; Taneja-Bageshwar *et al.*, 2009). We found a precursor for insect kinins in the *R. prolixus* genome, encoding for six kinins with the classical C terminal sequence, and five neuropeptides with the C-terminal pentapeptide FSX_1WA-NH_2 (X_1 = S, H). The kinin precursor encodes for 12 non-amidated peptides (see sequences in Table 2). An effect on the regulation of diuresis-related processes has been shown for the leucokinin of *R. prolixus* (Donini *et al.*, 2008; Te Brugge *et al.*, 2009). Mimetic analogues of insect kinins lead in the development of agents capable of disrupting kinin-regulated events as diuresis. Kinin analogues have been already suggested as tools for pest insect management (Nachman *et al.*, 2009; Taneja-Bageshwar *et al.*, 2009).

The knowledge of new and specific sequences for kinin neuropeptides from *R. prolixus* open a path for the investigation of new methods to control Chagas disease vectoring.

In *R. prolixus*, diuresis begins with the storage of the ingested blood into the anterior midgut. Ions and water move across the epithelium of the anterior midgut to the haemolymph, and from there on into the lumen of the Malpighian tubules. These processes are coordinated by a combination of neurohormones, neurotransmitters and neuromodulators. Interestingly, we found expression of neuroparsin A (NPA) precursor in ESTs derived from the anterior midgut, suggesting a role for NPA peptides in diuresis. An antidiuretic role of NPs has been previously suggested in *Locusta migratoria* and *Schistocerca gregaria* (Fournier *et al.*, 1994; Girardie *et al.*, 1998).

Development-related hormones. We identified three NPGs in the *R. prolixus* genome related to development and metamorphosis: allatostatin-B (AST-B), eclosion hormone (EH) and corazonin (CZ). AST-B precursor is very similar in sequence to that reported in *Bombyx mori*

(Hua *et al*., 1999). Interestingly, in addition to its expression in the nervous system, CZ has been discovered in testes-derived ESTs, as well as in the haemolymph (see below). Taking into account that the CZ receptor shares sequence similarity with the gonadotropin-releasing hormone receptor family (Kim *et al*., 2004), our findings support the hypothesis that an ancestral function of CZ is related to the control of reproduction.

Orcokinins. The orcokinins (OK) are a family of peptides that have been identified in crustaceans and in the cockroach *Blattella germanica* (Pascual *et al*., 2004), the locust *S. gregaria* (Hofer *et al*., 2005) and *Ap. mellifera* (Hummon *et al*., 2006). Putative homologues of OKs have been identified by database search in the genomes of *An. gambiae* (Pascual *et al*., 2004) and *D. melanogaster* (Liu *et al*., 2006), but expression has not yet been empirically confirmed. We identified the OK precursor in the *R. prolixus* genome, encoding three new insect OKs along with four putative peptides not related with OKs (Table 2). The expression of OK NPG was demonstrated by both RACE-PCR and MS. Moreover, we detected OKs in the haemolymph as well (see below).

SIF-amide peptides. SIF-amide is one of the most conserved neuropeptides in arthropods (Verleyen *et al*., 2009). Here we report a precursor encoding for a new isoform of this peptide family: TYKKPPFNGSIF-NH$_2$. We were able to confirm the presence of this peptide and three truncated forms by means of MS *de novo* sequencing (Table 2), as it has been observed in the *Ap. mellifera* peptidome (Hummon *et al*., 2006). SIF-amide has been detected in most insects with complete metamorphosis (Verleyen *et al*., 2009) and reported to be absent from Orthoptera after an exhaustive peptidomic search in *S. gregaria* and *L. migratoria* (Clynen *et al*., 2001, 2003). Our data indicate that the peptide is not exclusive to holometabolous insects. In order to determine whether the expression pattern is related to that in holometabolous insects, we studied the expression of SIF-amide in the nervous system of *R. prolixus*. Inmunohistochemistry reveals that SIF-amide expression is restricted to two pairs of large medial neurones in the *R. prolixus* brain, with projections extending along the entire length of the nerve cord, and with abundant varicosities along the central nervous system (Fig. 3). This expression pattern is similar to that observed in *D. melanogaster* and several other holometabolous insects (Verleyen *et al*., 2004; Terhzaz *et al*., 2007; Roller *et al*., 2008; Verleyen *et al*., 2009).

Conserved NPGs

Despite specific features and signatures at the sequence level, the structure of the pre-propeptides identified here is conserved in terms of the location of the immature peptides. For instance, the structure of the adipokinetic hormone (AKH) is conserved amongst insects, including *R. prolixus*, with the short neuropeptide sequence located immediately after the N-terminal signal peptide and tailed by a longer associated peptide. The multiple repeated peptide sequences separated by cleavage sites showed here are similar to other insect pre-propeptides, including tachykinins and allatostatin A (AST-A), both with 13 derived peptides (see sequences in Table 2).

Neuropeptide-like precursor 1 (NPLP1), a gene displaying all the hallmarks of a NPG, was first identified in *D. melanogaster* by peptidomic analysis (Baggerman *et al*., 2002), and later in other insects such as *Ap. mellifera* (Hummon *et al*., 2006), *B. mori* (Roller *et al*., 2008) and *T. castaneum* (Li *et al*., 2008). However, its physiological role is so far unknown. The NPLP1 gene in *R. prolixus* has slight similarity with its orthologue in *D. melanogaster* (Table 2) with 20 putative derived neuropeptides. By MS we identified seven of these mature peptides in the brain of *R. prolixus* (Table 2).

Partially identified precursors and missing neuropeptides in *R. prolixus*

We identified partial transcriptional units of some NPGs in *R. prolixus* contigs by a TBLASTN search. However, as a result of possible gaps in the genomic sequence, neither complete bioinformatic prediction nor cloning of RACE products could be carried out. These peptides include: pigment dispersing factor, insulin like peptides 1, 2, 3 and 6, *Apis*-like prohormone 2 and 3, diuretic hormone 44, ecdysis triggering hormone and AKH/corazonin-related peptide (ACP; Table S2). From these, a fragment of ACP from the *R. prolixus* genome sequence was reported while this paper was being revised (Hansen *et al*., 2010). Here we report a more extended fragment from this gene, including the signal peptide (Table S2).

Some neuropeptides that are conserved in other insects seem to be absent in the genome of *R. prolixus*, such as allatostatin-C (AST-C) and proctolin. We cannot rule out the possibility that gaps in the genome sequence available or highly diverging sequences were not detected by our homology search, and not yet validated by MS. However, we cannot exclude the presence of AST-C in *R. prolixus*, as a *bona fide* cognate receptor of AST-C was found by TBLASTN search ($E = 5e^{-108}$; M. Sterkel, unpubl. data), but we do not have evidence of any cognate receptor for the proctolin, suggesting the lack of proctolin in *R. prolixus*. In this respect, it is known that insects can either keep, duplicate or eliminate hormonal systems during evolution and conservation is not mandatory (see Table 1).
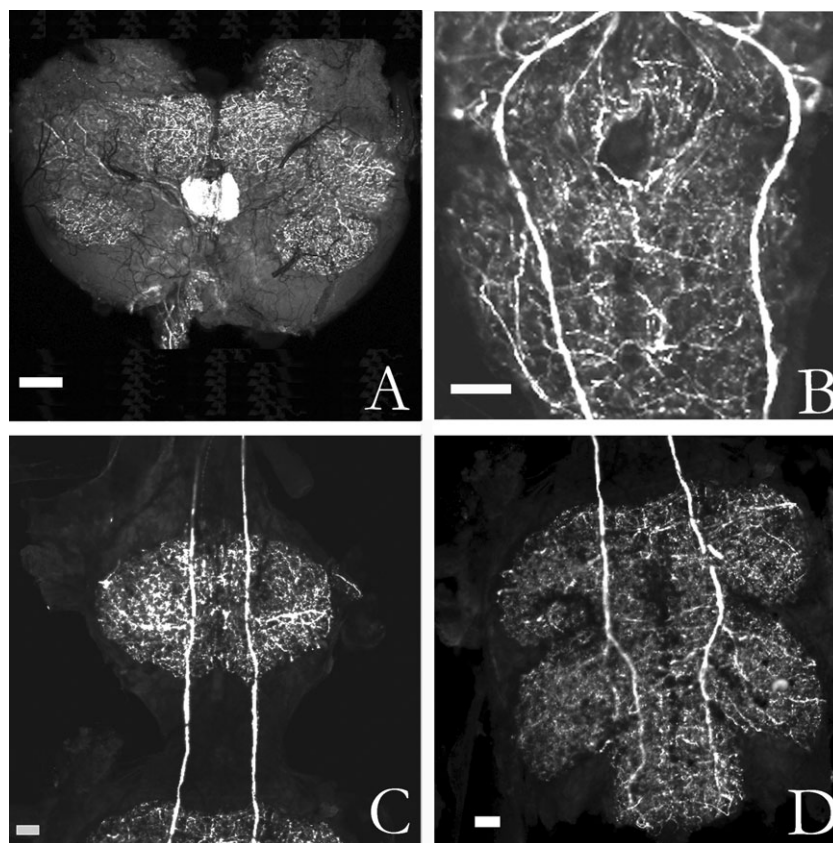
Figure 3. SIF-amide-like immunoreactivity (SLI) in the central nervous system of *Rhodnius prolixus*. (A) Dorsal view of the brain. Note the two pairs of brightly stained cells in the medial part of the brain, and the profuse neuropili. (B) View of the suboesophageal ganglion showing projection descending from the brain and profuse neuropili presenting SLI. (C) View of the prothoracic ganglion showing projection descending from the brain and profuse neuropili presenting SLI. (D) View of the mesothoracic ganglionic mass prothoracic ganglion showing projection descending from the brain and profuse neuropili presenting SLI. Scale bars = 100 μm.

Circulating neuropeptides in the haemolymph

After its identification in the nervous system, the hormonal function of a neuropeptide should be confirmed by its presence in circulating fluids. This proof is, however, omitted for most insect hormones because of the technical complexity associated with analysis of the haemolymph. The identification of highly diluted molecules such as neuropeptides is masked by the complex composition and high concentration of the proteins present in the haemolymph, a minor problem for the detection of highly concentrated peptides such as antimicrobial peptides. However, we performed a 'proof of concept' experiment – based on the molecular data that we have generated – to determine the neurohormonal function of some of the peptides. Using a protocol that includes nano-liquid chromatography separation and matrix assisted laser desorption ionization-time of flight (MALDI-TOF) MS/MS, we were able to detect by mass-match and MS/MS analysis two OKs, CZ, pyro-MSP and sNPF in *R. prolixus* haemolymph (Fig. 4). In the case of the OKs, for instance, it has been suggested that they have a neuroregulatory function in

the circadian clock of *Leucophaea maderae* (Hofer & Homberg, 2006); the presence of OKs in the haemolymph of *R. prolixus* indicates in addition a neurohormonal role. Other brain neuropeptides could also be present in the haemolymph, but might have remained undetectable to our experimental conditions as a result of concentrations below the limit of detection of our technique.

Concluding remarks

The combination of bioinformatics, genomics, transcriptomics and proteomics has produced relevant advances in neuroendocrinology. Nevertheless – and surprisingly – these combined studies had not been applied to insects of medical interest until very recently (Predel *et al.*, 2010). Here we developed a simple and straightforward workflow that yielded a comprehensive analysis of neuropeptides and neuropeptide precursors of an insect disease vector. From the mature peptides previously identified by MS in the *R. prolixus* brain (Ons *et al.*, 2009) and by sequence comparison we discovered 32 new genes encoding the neuropeptide precursors, verified the expression of their
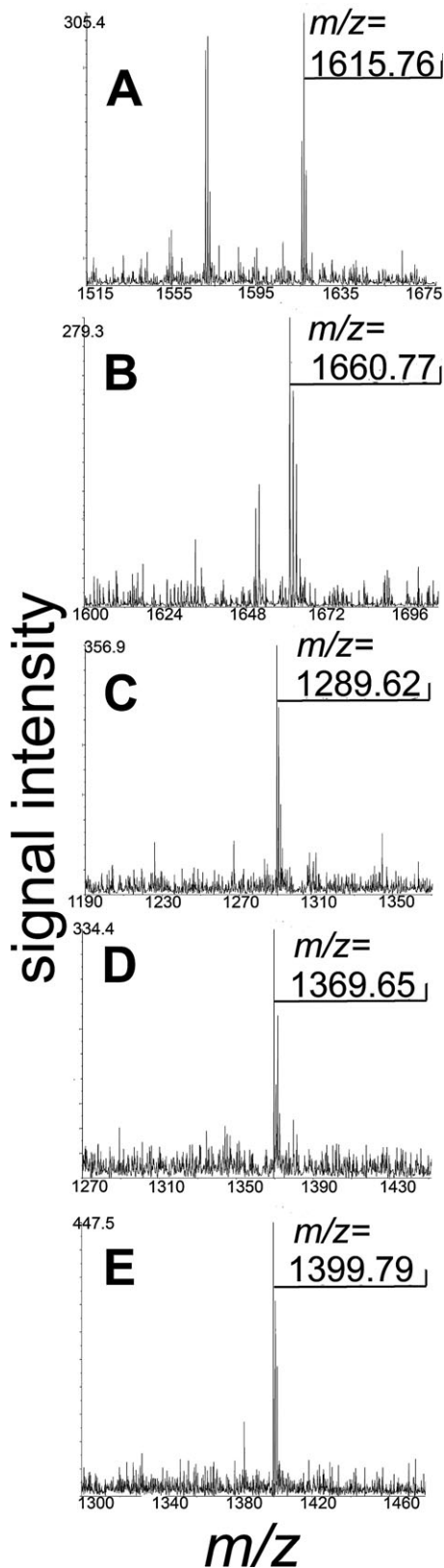
mRNA, deduced 194 new neuropeptides and extended to 82 the number of mature peptides that occurs in the brain. We also used the mass information to detect several neuropeptides in the circulating haemolymph.

Altogether, our results indicate that the neuropeptidome of *R. prolixus* is quite different from that of other insects with a sequenced and annotated genome, all of them holometabolous. However, broader comparative studies are required to determine whether these features are representative of the physiology of a hemimetabolous insect. Our study, which identified particular features in the peptidome of *R. prolixus*, provides a promising starting point for physiological and pharmacological studies aimed at the design of next-generation insecticides such as peptidomimetics, which are expected to be species-specific and environmentally friendly.

This work will also contribute to the annotation of neuropeptide-encoding genes in the genome of *R. prolixus*, which is in its final stage of assembly. We provide experimental evidence of the activity of several genes that validates the bioinformatic predictions. In addition, and more significantly, we provide a list of the neuropeptides derived from all the identified genes. The data presented here, including gene sequence and calculated and validated neuropeptide mass, provide a powerful tool for the insect research community. Functional studies involving RNA interference, tissue specific gene expression, quantitative proteomics, peptidome tissue mapping and MS imaging are now possible in *R. prolixus*.

Experimental procedures

Gene identification

Trace of *R. prolixus* whole genome sequence (WGS) and ESTs (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=TraceArchive&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch), EST databases from anterior midgut, posterior midgut, rectum, testes and whole organism (P. Oliveira, unpubl. results) and *R. prolixus* WGS assembly databases were used for homology and MS based searches. *R. prolixus* WGS data were produced at the Washington University School of Medicine in St. Louis, and can be obtained from http://genome.wustl.edu/genomes/view/rhodnius_prolixus/. Searches were performed by local TBLASTN by using BLOSUM62. Prediction of the gene structure and open reading frame was carried out with FGENESH, (http://www.softberry.com); GENSCAN (http://genes.mit.edu/GENSCAN.html), AUGUSTUS (http://augustus.gobics.de/) and by manual correction. Signal sequences were predicted with the SIGNALP server (Bendtsen *et al.*, 2004; http://www.cbs.dtu.dk/services/SignalP/). A statistical neuropeptide prediction algorithm was used on all putative neuropeptide genes to predict potential bioactive peptide sequences and molecular weights (http://www.neuroproteomics.scs.uiuc.edu/neuropred.html). Although these tools assist peptide prediction from prohormones, homology of a precursor between species also aids in neuropeptide prediction. Thus, after using the statistical approach, the prohormones

Figure 4. MALDI TOF-mass spectrometry of peptides identified in haemolymph corresponding to (A) orcokinin-1; (B) orcokinin-2; (C) pyro-myosuppressin; (D) corazonin; (E) short neuropeptide F.

with homology to known prohormones were manually examined to create the most likely set of mature peptides.

The transcript of each gene was confirmed by either reverse transcription PCR, RACE PCR and/or EST sequences. Template cDNA was prepared from the nervous system of adult *R. prolixus* at different times post-feeding. Primers used in PCR are listed in Table S1. RACE was performed using the GeneRacer Kit with SuperScript III RT and Zero Blunt® TOPO® PCR Cloning kit for Sequencing (Invitrogen, Buenos Aires, Argentina), according to the manufacturer's instructions.

Insect rearing and sample preparation

A colony of *R. prolixus* was maintained in our laboratory in a 12 h light/dark schedule at 28 °C. Insects were fed on chicken blood.

Brains from male and female adult insects were manually dissected and immediately placed in ice-chilled *R. prolixus* saline (NaCl, 129 mM; KCl, 8.6 mM; CaCl$_2$ 2.0 mM; MgCl$_2$ 8.5 mM; NaHCO$_3$ 10.2 mM; NaH$_2$PO4 4.3 mM; Hepes 8.6 mM, pH 7). For tissue extraction, 20 brains were placed in 100 μl methanol/water/acetic acid (90, 9, 1, v/v/v), sonicated for 5 min and centrifuged for 10 min at 7500 *g*. The supernatant was collected and the pellet was re-extracted twice. The collected supernatant was placed over vacuum to remove organic solvents, re-diluted in 20 μl 0.1% trifluoroacetic acid (TFA) and de-salted using a C18 extraction disc (Varian, Darmstadt, Germany) activated with 80% acetonitrile (ACN)/0.1% TFA. C18 extraction material was washed with 0.1% TFA, the sample was loaded and peptides were eluted with 70% ACN/0.1% TFA. Organic solvent was removed under vacuum and the sample re-diluted in 20 μl 0.1% TFA.

Haemolymph from six adult male and female insects was collected on an ice cold Eppendorf tube with a micropipette tip by cutting the second and third abdominal femurs (10 μl/insect). Haemolymph was pooled and mixed with an equal volume of 0.1% TFA. The sample was then boiled for 10 min and centrifuged for 10 min at 12 000 *g*. Three microlitres of the supernatant was de-salted with C18 material as described for brain tissue.

Nano C18 reversed-phase (RP) LC. Nano-RP LC separation was performed on a Dionex nano-LC system (LC Packings Idstein, Germany) equipped with autosampler, loading pump, nano pump, helium degasser, precolumn (25 × 0.15 mm) working in back-flush, analytical column (150 × 0.075 mm) and a multichannel detector. Precolumns were self-packed using Dr Maisch Reprosil-Pur 120 ODS-3 (5 μm particle size and 12 nm pore size, Dr Maisch, Ammerbuch-Entringen, Germany). Analytical columns were self-packed with C18 RP material (Vydac MS218, 5 μm particle size and 30 nm pore size,Vydac, Hesperia, CA, USA). Brain or haemolymph samples were loaded onto the pre-column at a flow rate of 10 μl/min in loading buffer (0.1% v/v TFA) in water for 10 min. After valve switching peptides were eluted by a linear gradient of 10–60% v/v solvent B (80% v/v ACN, 0.1% v/v TFA in water) over 180 min. The flow rate was 0.3 μl/min. The gradient was then raised from 60 to 100% of solvent B over 10 min. The eluate was mixed with α-cyano-4-hydroxicinamic acid (10 mg/ml in 70% ACN, 0.1% TFA) containing Glu-fibrinogen peptide (10 fmol/μl; Sigma, St Louis, MO, USA) as internal standard, and directly spotted on a steel MALDI target plate using a Probot microfraction collector (LC Packing). Matrix was delivered with a flow rate of 0.9 μl/min and fractions were spotted at intervals of 15 s onto stainless steel LC-MALDI plates (Applied Biosystems/MDS Sciex, Darmstadt, Ternary). All samples were processed in triplicate.

Mass spectrometry. MALDI-TOF- MS/MS was performed on an ABI 4800 analyzer (Applied Biosystems/MDSSciex). Each spot was first analysed by MS. Spectra were recorded between 600 and 4000 m/z. Job-wide interpretation of the MS data allowed the 15 peptides with the highest intensities (with signal to noise ratios ≥30) to be selected for sequencing by MS/MS. 1000 (3800 J) and 5000 (4500 J) laser shots were applied for MS and MS/MS, respectively. Collision energy was set to 1×10^{-6} Torr, with the potential difference between the source acceleration voltage and the collision cell set at 1 kV. An eight-point plate model calibration was performed with a Calibration Mixture 5 kit for Proteomics Analyzer (Applied Biosystems/MDX Sciex).

Peptide identification. MS/MS spectra were processed and peak lists were used for *de novo* sequencing. *De novo* sequencing was first performed automatically by PEAKS STUDIO 3.1 software (Bioinformatic Solutions Waterloo, Ontario, Canada) with the following settings: enzyme: none; variable modification; pyro-glutamic of glutamine (Q), C-terminal amidation; parent ion mass error tolerance = 0.1 Da; fragment ion mass error tolerance = 0.1 Da. Manual analysis was performed for every spectrum to confirm or modify auto *de novo* results and to detect post translational modifications as oxidation of Met, N-terminal pyro-glutamic from Gln, C-terminal amidation and N-terminal acetylation from Ala. All results reported here are from careful manual analysis of each spectrum, identifying accurate series of b-, y-, a- and z-ions and immonium ions confirming sequences.

Immunohistochemistry. The dorsal cuticle and tissues were removed from fifth instar *R. prolixus* nymphs under *R. prolixus* saline (NaCl, 129 mM; KCl, 8.6 mM; CaCl$_2$ 2.0 mM; MgCl$_2$ 8.5 mM; NaHCO$_3$ 10.2 mM; NaH$_2$PO4 4.3 mM; Hepes 8.6 mM, pH 7), exposing the brain and adjacent nervous system. Tissues were fixed *in situ* in 2% paraformaldehyde (pH 7) for 24 h at 4 °C. The nervous tissue was removed, washed with phosphate buffered saline (PBS) (pH 7), transferred into 4% Triton X-100 (Sigma, St. Louis, MO, USA) with 2% bovine serum albumin (BSA) and 10% normal goat serum (NGS) in PBS for 1 h at room temperature, and then washed several times with PBS at room temperature. Primary antiserum anti-SIFamide, kindly provided by Dr J. A. Veenstra (Bordeaux University, France), was pre-incubated at 1:2000 dilution in a 0.4% solution of Triton X-100 in PBS with 2% BSA and 2% NGS for 24 h at 4 °C prior to use. Tissues were incubated in primary antiserum for 48 h at 4 °C, and then washed for 18–24 h at 4 °C in PBS. Tissues were then incubated in Cy3-labelled goat antirabbit immunoglobulin solution (Jackson, West Grove, PA, USA) at 1:500 dilution in 10% NGS in PBS for 12 h at 4 °C, and finally rinsed several times in PBS at room temperature. Control experiments omitting the primary antiserum were also performed. Tissues were mounted on microscope slides in Mowiol (Sigma, St. Louis, MO, USA). Images were acquired using a confocal microscope LSM-510-META (Carl Zeiss, Jena, Germany) and processed with the software ZEISS LSM IMAGE BROWSER.

Acknowledgements

References

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.

Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**: 783–795.

Brugge, V.A., Schooley, D.A. and Orchard, I. (2008) Amino acid sequence and biological activity of a calcitonin-like diuretic hormone (DH31) from *Rhodnius prolixus*. *J Exp Biol* **211**: 382–390.

Clynen, E., Baggerman, G., Veelaert, D., Cerstiaens, A., Van der Horst, D., Harthoorn, L. *et al.* (2001) Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, *Locusta migratoria*. *Eur J Biochem* **268**: 1929–1939.

Clynen, E., Huybrechts, J., De Loof, A. and Schoofs, L. (2003) Mass spectrometric analysis of the perisympathetic organs in locusts: identification of novel periviscerokinins. *Biochem Biophys Res Commun* **300**: 422–428.

Dai, L., Zitnan, D. and Adams, M.E. (2007) Strategic expression of ion transport peptide gene products in central and peripheral neurons of insects. *J Comp Neurol* **500**: 353–367.

Donini, A., O'Donnell, M.J. and Orchard, I. (2008) Differential actions of diuretic factors on the Malpighian tubules of Rhodnius prolixus. *J Exp Biol* **211**: 42–48.

Fournier, B., Guerineau, N., Mollard, P. and Girardie, J. (1994) Effects of two neuronal antidiuretic molecules, neuroparsin

and 5-hydroxytryptamine, on cytosolic free calcium monitored with indo-1 in epithelial and muscular cells of the African locust rectum. *Biochim Biophys Acta* **1220**: 181–187.

Girardie, J., Huet, J.C., Atay-Kadiri, Z., Ettaouil, S., Delbecque, J.P., Fournier, B. *et al.* (1998) Isolation, sequence determination, physical and physiological characterization of the neuroparsins and ovary maturing parsins of *Schistocerca gregaria*. *Insect Biochem Mol Biol* **28**: 641–650.

Hansen, K., Stafflinger, E., Schneier, M., Hauser, F., Cazzamali, G., Williamson, M. *et al.* (2010) Discovery of a novel insect neuropeptide signaling system closely related to the insect Adipokinetic Hormone and Corazonin hormonal systems. *J Biol Chem* **285**: 10736–10747.

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

Hofer, S. and Homberg, U. (2006) Evidence for a role of orcokinin-related peptides in the circadian clock controlling locomotor activity of the cockroach *Leucophaea maderae*. *J Exp Biol* **209**: 2794–2803.

Hofer, S., Dircksen, H., Tollback, P. and Homberg, U. (2005) Novel insect orcokinins: characterization and neuronal distribution in the brains of selected dicondylian insects. *J Comp Neurol* **490**: 57–71.

Hua, Y.J., Tanaka, Y., Nakamura, K., Sakakibara, M., Nagata, S. and Kataoka, H. (1999) Identification of a prothoracicostatic peptide in the larval brain of the silkworm, *Bombyx mori*. *J Biol Chem* **274**: 31169–31173.

Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A. *et al.* (2006) From the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* **314**: 647–649.

Kim, Y.J., Spalovska-Valachova, I., Cho, K.H., Zitnanova, I., Park, Y., Adams, M.E. *et al.* (2004) Corazonin receptor signaling in ecdysis initiation. *Proc Natl Acad Sci USA* **101**: 6704–6709.

Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G. *et al.* (2008) Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. *Genome Res* **18**: 113–122.

Liu, F., Baggerman, G., D'Hertog, W., Verleyen, P., Schoofs, L. and Wets, G. (2006) In silico identification of new secretory peptide genes in *Drosophila melanogaster*. *Mol Cell Proteomics* **5**: 510–522.

Nachman, R.J., Pietrantonio, P.V. and Coast, G.M. (2009) Toward the development of novel pest management agents based upon insect kinin neuropeptide analogues. *Ann N Y Acad Sci* **1163**: 251–261.

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

Neupert, S., Russell, W.K., Predel, R., Russell, D.H., Strey, O.F., Teel, P.D. *et al.* (2009) The neuropeptidomics of *Ixodes scapularis* synganglion. *J Proteomics* **72**: 1040–1045.

Neupert, S., Russell, W.K., Russell, D. H. and Predel, R. (2010) Two capa-genes are expressed in the neuroendocrine system of *Rhodnius prolixus*. *Peptides* **31**: 408–411.

Ons, S., Richter, F., Urlaub, H. and Pomar, R.R. (2009) The neuropeptidome of *Rhodnius prolixus* brain. *Proteomics* **9**: 788–792.

Orchard, I. and Lange, A. (2006) Insect myosuppressins/ FMRFamides and FL/IRFamides/NPFs. In *Handbook of Biologically Active Peptides* (Kastin, A., ed.), pp. 193–199. Elsevier-Academic Press, London.

Paluzzi, J.P. and Orchard, I. (2010) A second gene encodes the anti-diuretic hormone in the insect, *Rhodnius prolixus*. *Mol Cell Endocrinol* **317**: 53–63.

Paluzzi, J.-P., Russell, W.K., Nachman, R.J. and Orchard, I. (2008) Isolation, cloning and expression mapping of a gene encoding an antidiuretic hormone and other CAPA-related peptides in the disease vector, *Rhodnius prolixus*. *Endocrinology* **149**: 4638–4646.

Pascual, N., Castresana, J., Valero, M.L., Andreu, D. and Belles, X. (2004) Orcokinins in insects and other invertebrates. *Insect Biochem Mol Biol* **34**: 1141–1146.

Picollo, M.I., Vassena, C., Santo Orihuela, P., Barrios, S., Zaidemberg, M. and Zerba, E. (2005) High resistance to pyrethroid insecticides associated with ineffective field treatments in *Triatoma infestans* (Hemiptera: Reduviidae) from Northern Argentina. *J Med Entomol* **42**: 637–642.

Predel, R., Russell, W.K., Russell, D.H., Lopez, J., Esquivel, J. and Nachman, R. (2008) Comparative peptidomics of four related hemipteran species: pyrokinins, myosuppressin, corazonin, adipokinetic hormone, sNPF, and periviscerokinins. *Peptides* **29**: 162–167.

Predel, R., Neupert, S., Garczynski, S., Crim, J., Brown, M., Russell, W.K. *et al.* (2010) Neuropeptidomics of the mosquito *Aedes aegypti*. *J Proteome Res* **9**: 2006–2015.

Reumer, A., Van Loy, T., Clynen, E. and Schoofs, L. (2008) How functional genomics and genetics complements insect endocrinology. *Gen Comp Endocrinol* **155**: 22–30.

Riehle, M., Garczynski, S., Crim, J., Hill, C. and Brown, M. (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* **298**: 172–175.

Roller, L., Yamanaka, N., Watanabe, K., Daubnerova, I., Zitnan, D., Kataoka, H. *et al.* (2008) The unique evolution of neuropeptide genes in the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* **38**: 1147–1157.

Scherkenbeck, J. and Zdobinsky, T. (2009) Insect neuropeptides: Structures, chemical modifications and potential for insect control. *Bioorg Med Chem* **17**: 4071–4084.

Schoofs, L. and Nachman, R. (2006) Sulfakinins. In *Handbook of Biologically Active Peptides* (Kastin, A., ed.), pp. 183–187. Elsevier-Academic Press, London.

Taneja-Bageshwar, S., Strey, A., Isaac, R.E., Coast, G.M., Zubrzak, P., Pietrantonio, P.V. *et al.* (2009) Biostable agonists that match or exceed activity of native insect kinins on recombinant arthropod GPCRs. *Gen Comp Endocrinol* **162**: 122–128.

Te Brugge, V., Ianowski, J.P. and Orchard, I. (2009) Biological activity of diuretic factors on the anterior midgut of the blood-feeding bug, *Rhodnius prolixus*. *Gen Comp Endocrinol* **162**: 105–112.

Terhzaz, S., Rosay, P., Goodwin, S.F. and Veenstra, J.A. (2007) The neuropeptide SIFamide modulates sexual behavior in *Drosophila*. *Biochem Biophys Res Commun* **352**: 305–310.

Verleyen, P., Huybrechts, J., Baggerman, G., Van Lommel, A., De Loof, A. and Schoofs, L. (2004) SIFamide is a highly conserved neuropeptide: a comparative study in different insect species. *Biochem Biophys Res Commun* **320**: 334–341.

Verleyen, P., Huybrechts, J. and Schoofs, L. (2009) SIFamide illustrates the rapid evolution in Arthropod neuropeptide research. *Gen Comp Endocrinol* **162**: 27–35.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/j.1365-2583.2010.01050.x

Table S1. List of primers used for rapid amplification of cDNA ends to confirm the sequence and transcription of predicted neuropeptide precursor genes.

Table S2. Genes partially identified.

Please note: Neither the Editors nor Wiley-Blackwell are responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.