Insect Biochemistry and Molecular Biology xxx (2015) 1–17

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis

Sheila Ons [a, 1], Andrés Lavore [b, 1], Marcos Sterkel [c], Juan Pedro Wulff [a], Ivana Sierra [a], Jesús Martínez-Barnetche [d], Mario Henry Rodriguez [d], Rolando Rivera-Pomar [a, b, *]

[a] *Laboratorio de Genética y Genómica Funcional, Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Bvd 120 y 62, La Plata, Buenos Aires, Argentina*

[b] *Departamento de Ciencias Básicas y Experimentales, Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Monteagudo 2772, 2700 Pergamino, Buenos Aires, Argentina*

[c] *Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho, 373, bloco D. Prédio do CCS, Ilha do Fundão, Rio de Janeiro 21941-902, Brazil*

[d] *Centro de Investigaciones sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, Av. Universidad 655, Sta. María Ahuacatitlán, Cuernavaca, Mexico*

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form
29 April 2015
Accepted 1 May 2015
Available online xxx

Keywords:
Chagas disease
Insect
Triatoma
Transcriptome
Neuropeptides
GPCRs

ABSTRACT

The importance of Chagas disease motivated the scientific effort to obtain the complete genomic sequence of the vector species *Rhodnius prolixus*, this information is also relevant to the understanding of triatomine biology in general. The central nervous system is the key regulator of insect physiology and behavior. Neurohormones (neuropeptides and biogenic amines) are the chemical messengers involved in the regulation and integration of neuroendocrine signals. In insects, this signaling is mainly mediated by the interaction of neurohormone ligands with G protein coupled receptors (GPCRs). The recently sequenced *R. prolixus* genome provides us with the opportunity to analyze this important family of genes in triatomines, supplying relevant information for further functional studies. Next-generation sequencing methods offer an excellent opportunity for transcriptomic exploration in key organs and tissues in the presence of a reference genome as well as when a reference genome is not available. We undertook a genomic analysis to obtain a genome-wide inventory of opsines and the GPCRs for neurohormones in *R. prolixus*. Furthermore, we performed a transcriptomic analysis of *R. prolixus* central nervous system, focusing on neuropeptide precursor genes and neurohormone and opsines GPCRs. In addition, we mined the whole transcriptomes of *Triatoma dimidiata*, *Triatoma infestans* and *Triatoma pallidipennis* – three sanitary relevant triatomine species – to identify neuropeptide precursors and GPCRs genes. Our study reveals a high degree of sequence conservation in the molecular components of the neuroendocrine system of triatomines.

1. Introduction

Chagas disease is an important but neglected human disease, with 8 million people infected in Latin America and a fifth of the population of the region at risk; 30–40% of the infected people develop cardiomyopathy and/or digestive syndromes (Rassi et al., 2010.). The causative agent, the protozoan *Trypanosoma cruzi*, is mostly transmitted to humans by the feces of triatomine bugs. From these, the three most important vector species are *Triatoma infestans* in southern South America, *Rhodnius prolixus* in northern South America and Central America, and *Triatoma dimidiata* in

* Corresponding author. Laboratorio de Genética y Genómica Funcional, Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Bvd 120 y 62, La Plata, Buenos Aires, Argentina.

E-mail addresses: sheila.ons@presi.unlp.edu.ar (S. Ons), andreslavore@unnoba.edu.ar (A. Lavore), msterkel29@googlemail.com (M. Sterkel), juanpwulff@hotmail.com (J.P. Wulff), sierra.ivana@gmail.com (I. Sierra), jmbarnet@insp.mx (J. Martínez-Barnetche), mhenry@insp.mx (M.H. Rodriguez), rrivera@unnoba.edu.ar (R. Rivera-Pomar).

[1] These authors equally contributed to the present work.

http://dx.doi.org/10.1016/j.ibmb.2015.05.003
0965-1748/

northern South America, Central America and Mexico (Rassi et al., 2010). The importance of Chagas disease motivated the scientific effort to obtain the complete sequence of *R. prolixus* genome. This information will be highly relevant for our understanding of triatomines biology.

The central nervous system (CNS) is the key regulator of insect physiology and behavior. Sensory imputes are integrated in CNS to generate behavioral responses and regulation of physiological processes such as feeding, development and reproduction. Neuropeptides and biogenic amines are the chemical messengers involved in the regulation and integration of neuroendocrine signals. In insects, this signaling is mainly mediated by the interaction of neuropeptide and neurohormone ligands with G protein coupled receptors (GPCRs), a large superfamily of proteins characterized by the presence of seven α-helical transmembrane domains. Biogenic amines and opsins GPCRs belong to family A (or rhodopsin-like), whereas neuropeptide GPCRs belong to family A and family B (or calcitonin-like) (Hauser et al., 2008; Li et al., 2013).

The insect neuroendocrine system is a promising target for a novel generation of insecticides that offer improved selectivity and environmental compatibility. Manipulations in the interaction of neuropeptides and their receptors through peptidomimetics or pseudopeptides could disrupt fundamental physiological processes. These facts motivated our comprehensive analysis of *R. prolixus* neuropeptide complement over the last few years (Ons et al., 2011, 2009; Sterkel et al., 2012). Since functional information regarding individual genes encoding *R. prolixus* GPCRs has arisen recently (Paluzzi et al., 2010; Paluzzi and O'Donnell, 2012; Paluzzi et al., 2014; Zandawala et al., 2013), a genome-wide comprehensive analysis of the GPCRs of *R. prolixus* has not been performed. The information provided here will boost the functional analysis on the neuroendocrine system in triatomines.

Next-generation sequencing (NGS) methods (Roche 454, Solexa/Illumina, etc) offer an excellent opportunity for transcriptomic exploration in key organs and tissues in the presence of a reference genome as well as when a reference genome is not available. However, with a few exceptions (Wheeler et al., 2013; Zhang et al., 2012), NGS methods have not been extensively used for a comprehensive analysis of insect nervous system to date.

Here we performed a genome-wide identification and annotation of *R. prolixus* GPCRs for neurohormones and opsines, using query orthologues sequences from *Anopheles gambiae*, *Apis mellifera*, *Bombyx mori*, *Drosophila melanogaster* and *Tribolium castaneum*. The results allowed us to make comparisons of neuroendocrine systems in different insect species, and to provide relevant information for further functional studies in *R. prolixus*. Furthermore, in this study we undertook an NGS approach using HiSeq2000 (Illumina) platform to analyze the transcriptome of *R. prolixus* CNS, focusing on neuropeptide precursor genes, neurohormone and opsines GPCRs. We also mined the whole transcriptomes of the other three sanitary relevant triatomine species (*T. dimidiata*, *T. infestans* and *Triatoma pallidipennis*) to identify neuropeptide precursor and GPCRs ESTs. Our genomic and transcriptomic analysis reveals a high degree of conservation in the neuroendocrine system in triatomines.

2. Material and methods

2.1. Identification of R. prolixus GPCRs

Using *D. melanogaster* (Hewes and Taghert, 2001), *A. gambiae* (Hill et al., 2002), *A. mellifera* (Hauser et al., 2006), *B. mori* (Fan et al., 2010) and *T. castaneum* (Hauser et al., 2008) sequences as references, TBLASTN searches were performed for GPCR candidates in the transcripts (RproC1.1) and genomic scaffolds (RproC1) databases of *R. prolixus* (http://www.vectorbase.org). The full sequences of candidate proteins were identified and/or reconstructed from the databases by manual analysis. 7TM domains and phylogenetic analysis methods were adopted as criteria for this study. Structural information for every candidate GPCRs was predicted with the server TMHMM (v2.0) from the Centre for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/TMHMM/).

2.2. Phylogenetic analysis

We compared the neurohormone and opsin GPCRs of *R. prolixus* and *Triatoma* spp. with the collections of orthologues GPCRs previously published for *D. melanogaster*, *T. castaneum*, *B. mori*, *A. mellifera* and *A. gambiae*. Gene bank accession numbers were used for the identification of sequences.

Phylogenies of GPCRs were based on sequence alignments using CLUSTAL W (Larkin et al., 2007). These alignments were later used for phylogenetic analysis employing Bayesian inference and the BEAST program (Drummond and Rambaut, 2007). The parameters used for the analysis were: Number of substitution types: 6 (GTR), Substitution model: Blosum62, Number of generations: 200.000 and a Sample tree every 10 generations. The trees were drawn using Figtree software (http://tree.bio.ed.ac.uk/software/figtree/), burning the first five trees.

2.3. Insect rearing

Colonies of *R. prolixus*, *T. dimidiata*, *T. infestans* and *T. pallidipennis* from Centro de Referencia de Vectores (CeReVe), Córdoba, Argentina were maintained in Centro Regional de Estudios Genomicos at 28 °C and a partial humidity of 70% with a 12 h light/dark schedule. Insects were regularly fed on chickens.

2.4. R. prolixus CNS transcriptome preparation and sequencing

CNSs were microdissected from adult male and female *R. prolixus* at different times post-blood meal (PBM): unfed, 1 h PBM, 4 h PBM and 24 h PBM. Total RNA was isolated using Trizol (Ambion, Sao Paulo, Brazil). Four cDNA libraries were independently constructed with 1 μg of RNA from each group using Mint-2 Kit (Evrogen, Moscow, Russia), according to the manufacturer instructions for Illumina RNAseq.

The cDNA libraries produced (unfed, 1 h, 4 h and 24 h PBM) were independently sequenced using the HiSeq 2000 platform (Illumina) at the Goettingen University sequencing facility (Germany). The four libraries data was joined in a unique data set which was equally representative in number of reads for each sample.

2.5. Data filtering and assembly

Low quality sequences were filtered from the row data using the program package FASTX Toolkit to remove reads with quality scores lower than 20 and adaptors sequences. *R. prolixus* CNS dataset was assembled with Velvet (Zerbino and Birney, 2008) and Oases (Schulz et al., 2012) software packages using Kmer values between 31 and 43. The resulting contigs were merged with Velvet and Oases to keep a single and non redundant assembly. We filtered the loci with a confidence value of 1.0 in the final assembly.

2.6. Mapping to R. prolixus genome and differential expression analysis

In order to estimate the proportion of transcripts present in our transcriptome, we used the Bowtie2 software (Langmead and Salzberg, 2012) with default parameters to align the CNS *de novo*

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

assembly with the *R. prolixus* genome scaffolds (http://www.vectorbase.org). Furthermore, we mapped the *R. prolixus* CNS transcripts to *T. dimidiata* mitochondrial genome (Dotson and Beard, 2001) to evaluate the degree of coverage.

The raw read counts for each of the four Illumina libraries (four PBM conditions) were used as input to the DESeq R package (Anders and Huber, 2010) to perform pairwise differential expression analysis between the starved (basal) and fed insects. The estimate Size Factors function of the DESeq package was used to normalize gene counts.

2.7. Proteomes comparison and transcript annotation

The assembled dataset was used to identify the proportion of the core eukaryotic genome. We used Hidden Markov Model (HMM) profiles for 458 core eukaryotic proteins as provided by the CEGMA algorithm (Parra et al., 2007) and HMMER3 searches with the hmmscan command and the -T 40 and –domT 40 filters.

On the other hand, the CNS transcriptome assembly was used to find orthologous proteins in the following arthropods proteomes: *R. prolixus* (RproC1.2, www.vectorbase.org/organisms/rhodnius-prolixus/cdc/rproc1), *D. melanogaster* (UNIPROT-Swisprot, Adams et al., 2000), *Acyrthosiphon pisum* (translated genomic scaffolds, International Aphid Genomics Consortium, 2010), *I. scapularis* (IscaW1.2, www.vectorbase.org/organisms/ixodes-scapularis/wikel/iscaw1), *P. humanus* (PhumU1.2, Kirkness et al., 2010), *L. longipalpis* (LlonJ1.0, www.vectorbase.org/organisms/lutzomyia-longipalpis/jacobina/llonj1), *P. papatasi* (PpapI1.0, www.vectorbase.org/organisms/phlebotomus-papatasi/israel/ppapi1), *C. quinquefaciatus* (CpipJ1.3, Arensburger et al., 2010), *A. gambiae* (AgamP4.2, Holt et al., 2002), *Aedes aegypti* (AaegL3.2, Nene et al., 2007), *A. albimanus* (AalbS1.2, Martines-Barnetche et al., 2012) and *Schistocerca gregaria* (translated genomic scaffolds, Badisco et al., 2011). The searches were done using BLASTX with a cut-off e-value of 0.0001, keeping only the best hit for each query. Furthermore, we performed bidirectional BLASTX between the CNS transcriptome of *R. prolixus* and *S. gregaria* to find the best reciprocal hits within these two insect species.

The functional annotation of *R. prolixus* neurotranscriptome was performed using Blast2Go software package (Gotz et al., 2008). The gene ontology annotations were made using the assembled transcript sequences that mapped to *R. prolixus* genome as input dataset. With *R. prolixus* CNS transcriptome completely annotated, using Fisher's Exact Test, we carried out an enrichment analysis in which our annotated dataset was used as test group and the annotated proteome of *R. prolixus* as reference group.

2.8. Blast search

We performed iterative searches to identify neuropeptides and GPCRs in the transcriptomes of *R. prolixus* CNS, *T. dimidiata*, *T. infestans* and *T. pallidipennis* (Lavore et al., 2014, unpublished). For neuropeptides, we used the specific sequences of *R. prolixus* described earlier as queries (Ons et al., 2011; Orchard et al., 2011; Sterkel et al., 2012; Zandawala et al., 2010) and orthologues from *B. mori*, *D. melanogaster* and *T. castaneum* for molecules not reported in *R. prolixus*. For GPCRs we used the sequences predicted in this study (see 2.1) and previous reports (Paluzzi and O'Donnell, 2012; Paluzzi et al., 2014; Zandawala et al., 2013) as queries. Orthologue GPCRs from *A. gambiae*, *B. mori*, *D. melanogaster*, *A. mellifera* and *T. castaneum* were used for those molecules that were not identified in *R. prolixus* databases. The searches were performed using local BLAST (Altschul et al., 1990) and a minimum *e*-value of 0.0001.

2.9. Structure analysis

For the identification of Signal peptide we used SignalP 3 (Bendtsen et al., 2004). For the prediction of convertase cleavage sites (Veenstra, 2000) the rules described for insect neuropeptide precursors were followed.

3. Results and discussion

3.1. GPCRs for neurohormones and opsines

We annotated in *R. prolixus* the neuropeptides, biogenic amines and opsines GPCRs genes using *D. melanogaster*, *A. mellifera*, *T. castaneum*, *B. mori* and *A. gambiae* orthologous as queries to datamine the genome of *R. prolixus*. Manual curation for the automated predictions is presented in Supplementary information 1. With this strategy we identified 70 putative GPCRs from *R. prolixus*, 62 of which represent predicted proteins that are reported here for the first time. 29 of the identified proteins had matching EST evidence in *R. prolixus* CNS, 36 in *T. dimidiata*, 34 in *T. infestans* and 40 in *T. pallidipennis* transcriptomes (Tables 1–4).

3.1.1. Biogenic amines

The already known biogenic amines – ligands of GPCRs in insects – are acetylcholine, dopamine, serotonin, octopamine and tyramine. These amines exert their action through 21 GPRCs in *D. melanogaster* (Hauser et al., 2006), 20 GPRCs in *T. castaneum* (Hauser et al., 2008) and 18 GPCRs in *A. pisum* (Li et al., 2013). We found 18 biogenic amine GPCRs in *R. prolixus* for the first time; 11 of these have ESTs evidence in *R. prolixus* CNS (Table 1, Fig. 1). We also found ESTs with strong match to *R. prolixus* biogenic amines in *T. dimidiata* (13 ESTs), *T. infestans* (11 ESTs) and *T. pallidipennis* (13 ESTs) transcriptomes.

Basing on phylogenetic analysis and sequence similarity (Table 1, Fig. 1, Supplementary information 1) RPRC010907, RPRC001750 + RPRC001751 + RPRC007566 seem to be the orthologous of acetylcholine receptors; RPRC000473, RPRC014093, RPRC011175, RPRC013708 are related to dopamine receptors; RPRC014610 + RPRC001054, RPRC001507 + RPRC005349, RPRC011545, RPRC001341 are classified as octopamine receptors and RPRC008712 is close to octopamine/tyramine receptors. Five receptors are serotonin-like GPCRs in *R. prolixus* genome: RPRC008923, RPRC007788 + RPRC001792, RPRC010656, RPRC005858 + RPRC001892, RPRC010931. The analysis also enabled the classification of several *Triatoma* spp. sequences: TINF_IAZY42G02FR3G9, TDIM_isotig08113 and TINF_IAZY42G02H3MDU are serotonin receptors; TPAL H9TUR5Q02HGZGI, TPALH9TUR5Q02H2VUJ, TINF_isotig09594, TDIM_H9TUR5Q01BHBS4, TPAL_H9TUR5Q02G8UYN, TPAL_H9TUR5Q02GWI8U are dopamine receptors.

3.1.2. Neuropeptide GPCRs

3.1.2.1. Family A GPCRs. Forty-one family A GPCRs for neuropeptides and peptide hormones were identified in *R. prolixus*; 36 of these are reported here for the first time. Seventeen GPCRs have EST evidence in the CNS (Table 2). Furthermore, we found evidence for these GPRCs in *T. dimidata* (23 contigs), *T. infestans* (21) and *T. pallidipennis* (21).

Ligands of GPCRs were deduced by phylogenetic analysis and sequence similarity; basing on these, they were divided into 25 groups (Table 2, Fig. 2): ACP (RPRC000057 + RPRC004783), Allatostatin A (ASTA) (the paralogues RPRC000706, RPRC000708), Allatostatin C (ASTC) (RPRC013486), Allatotropin (AT) (KF740716), Bursicon (RPRC001663), CAPA peptides (ADG27752, ADG27753), CCHa (RPRC007766, RPRC000608), Crustacean Cardioactive

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

Table 1
Biogenic amine G protein coupled receptors identified in *Rhodnius prolixus* genomic information, CNS transcriptome and transcriptomes of Triatoma spp.

R. prolixus receptor number	Probable endogenous ligand	*Tribolium castaneum*	*Drosophila melanogaster*	*Anopheles gambiae*	*Apis mellifera*	*Bombix mori*	*Triatoma dimidiata*	*Triatoma infestans*	*Triatoma pallidipennis*	*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
Biogenic amine receptors													
RPRC010907	Acetylcholine	XP_008190280	AGE13747	XP_003436890	XP_006558421	XP_004922906	–	–	–	–	–	–	–
RPRC001750 + RPRC001751 + RPRC007566	Acetylcholine	XP_008200902	AAA28676	XP_314486	XP_395760	XP_004925506	2.00E−06	–	–	2.00E−09	0.38 (0.64)	1.60 (0.75)	1.13 (0.94)
RPRC014093	Dopamine	XP_008197846	AAA85716	XP_315207	NP_001011595	NP_001108459	8.00E−109	4E−48	1.00E−117	2.00E−09	0.38 (0.64)	1.60 (0.75)	1.13 (0.94)
RPRC011175	Dopamine	XM_008193803	AY150864	XP_003436820	XM_006561506	XM_004925908	–	–	–	–	–	–	–
RPRC013708	Dopamine	XP_008201093	NP_733299	XP_311193	XP_006617175	NP_001108338	2.00E−06	–	9.00E−08	1.00E−11	0.38 (0.64)	1.60 (0.75)	1.13 (0.94)
RPRC000473	Dopamine	NP_001280515	NP_001014757	XP_003436820	XP_006561570	XP_004925965	2.00E−06	4.00E−09	9.00E−09	3.00E−19	0.69 (0.94)		
RPRC014610 + RPRC001054	Octopamine	XP_008198079	NP_001034043	XP_003436281	XP_006557730	NP_001171666	2E−18	6.00E−10	2.00E−18	5.00E−23	0.38 (0.64)	1.60 (0.75)	1.13 (0.94)
RPRC001507 + RPRC005349	Octopamine	XP_974265	NP_651057	XP_003436280	XP_397139	XP_004922133	3.00E−19	3.00E−08	6.00E−19	9.00E−31	0.38 (0.64)	1.60 (0.75)	1.13 (0.94)
RPRC011545	Octopamine	NP_001280501	CAI56430	XP_003436280	XP_396348	NP_001171666	3.00E−18	1E−12	7.00E−19	–	–	–	–
RPRC001341	Octopamine	XP_008198470	NP_732541	XP_311113	XP_392093	NP_001091748	–	–	5.00E−06	9.00E−25	5.86 (0.59)	0.87 (0.99)	0.96 (1)
RPRC008712	Octopamine/ Tyramine	XP_970290	NP_524419	XP_312420	XP_393187	BAD11157	7.00E−08	2.00E−06	7.00E−13	2.00E−44	5.13 (0,38)	5.55 (0.30)	2.93 (0.52)
RPRC008923	Serotonin	XP_008197542	NP_001163201	XP_317820	NP_001164579	NP_001037502	1.00E−06	2.00E−43	6.00E−09	4.00E−15	0 (1)	0 (1)	0 (0.79)
RPRC007788 + RPRC001792	Serotonin	NM_001293606	M55533	XP_313129	NM_001077821	XM_004922054	3.00E−06	4.00E−08	2.00E−09	5.00E−12	0.38 (0.64)	1.61 (0.75)	1.13 (0.94)
RPRC010656	Serotonin	XP_008190659	NP_572358	XP_310742	–	XP_004924179	3.00E−105	–	–	–	–	–	–
RPRC005858 + RPRC001892	Serotonin	XP_966377	AAL28587	XP_307953	XP_006561625	XP_004923088	–	3E−25	2.00E−07	–	–	–	–
RPRC010931	Serotonin	XP_008197967	AAY84887	XP_308623	NP_001164579	XP_004925175	2.00E−07	7E−16	1.00E−08	6E−25	0 (1)	0 (1)	0 (0.78)

The three last columns indicate fold change and (p) in pairwise comparisons between unfed (basal) and different time points PBM in *R. prolixus* CNS transcriptomes.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

Table 2
Family-A G protein coupled receptors for neuropeptides identified in *Rhodnius prolixus* genomic information, CNS transcriptome and transcriptomes of Triatoma spp.

R. prolixus receptor number	Probable endogenous ligand	*Tribolium castaneum*	*Drosophila melanogaster*	*Anopheles gambiae*	*Apis mellifera*	*Bombix mori*	*Triatoma dimidiata*	*Triatoma infestans*	*Triatoma pallidipennis*	*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
							E-value	*E*-value	*E*-value	*E*-value			
RPRC000057 + RPRC004783	ACP	EEZ99294	AAC61523	XP_321591	NP_001035354	NP_001037049	3.00E−06	7E−19	6.00E−09	3.00E−27	0.4 (0.07)	0.82 (0.92)	0.56 (0.69)
RPRC004706	Allatostatin A	–	NP_524700	XP_003435928	XP_006560262	NP_001037035	7.00E−06	2.00E−14	–	9.00E−07	inf (0.005)	inf (0.003)	inf (0.029)
RPRC004708	Allatostatin A	–	NP_524700	XP_003435928	XP_006560262	NP_001037035	4.00E−06	3.00E−15	–	5.00E−07	inf (0.005)	inf (0.003)	inf (0.029)
RPRC013486	Allatostatin C	NP_001280521	AAL02125	XP_320274	XP_006560939	NP_001127736	–	2.00E−19	–	1.00E−54	0.76 (0.92)	0.48 (0.64)	1.09 (0.96)
KF740716	Allatotropin	XP_008199736	–	–	XP_006567285	NP_001127714	6.00E−08	1.00E−64	9.00E−87	–	–	–	–
RPRC001663	Bursicon	XP_008192239	AAF66608	XP_317111	XP_395206	XP_004923855	–	–	–	–	–	–	–
ADG27752	CAPA peptides	NM_001293616	AF522193	AY900217	NM_001098232	NM_001134253	3.00E−45	2E−15	2.00E−31	3.00E−07	∞(0.005)	∞ (0.004)	∞ (0.095)
ADG27753	CAPA peptides	NM_001293616	AF522193	AY900217	NM_001098232	NM_001134253	5E−30	2E−15	2.00E−31	2.00E−07	∞ (0.005)	∞ (0.004)	∞ (0.095)
RPRC007766	CCHa	XP_008199736	–	–	XP_006567285	NP_001127714	–	–	–	–	–	–	–
RPRC000608	CCHa	EFA09272	NP_611241	XP_313395	XP_006560492	NP_001127712	4.00E−09	5.00E−09	5.00E−08	2.00E−08	∞ (0.005)	inf (0.003)	∞ (0.029)
RPRC001248	Crustacean Cardioactive Peptide	XP_008192932	AAN10041	XP_321100	XP_001122652	NP_001127746	1E−60	–	9.00E−08	–	–	–	–
RPRC000969 + RPRC012063	Crustacean Cardioactive Peptide	NM_001083327	ACQ66048	XP_001238433	XP_003691184	NP_001127746	6E−25	6.00E−08	–	–	–	–	–
RPRC000523	Corazonin	–	NP_648571	XP_321555	NP_001137393	NP_001127719	–	5E−45	2.00E−08	–	–	–	–
RPRC008652	Ecdysis Triggering Hormone	NP_001076792	NP_996255	XP_003436278	XP_006570147	NP_001127741	–	–	–	–	–	–	–
RPRC000848	Ecdysis Triggering Hormone	NP_001076792	NP_650960	XP_312031	XP_006570147	NP_001165737	–	–	–	–	–	–	–
RPRC001551	FaLP	NP_001280540	NP_647758	XP_321207	ACI90286	NP_001037007	–	–	–	–	–	–	–
RPRC015267	FaLP/Proctolin	XP_008190951	NP_572183	XP_314133	–	–	1E−15	–	–	–	–	–	–
RPRC001428	GPA2/GPB5	XP_008191741	AAP45004	XP_557643	XP_006570731	NP_001127716	8.00E−07	–	–	–	–	–	–
RPRC00494	Kinin	XP_970102	NP_647968	XP_001230662	XP_006567021	NP_001127721	8.00E−07	1.00E−06	3.00E−10	3.00E−12	5.27 (0.36)	5.4 (0.2)	1.08 (0.95)
GL563029[a]	Long Neuropeptide F	XP_967232	AAP69822	XP_001688310	–	–	8.00E−09	1.00E−07	4.00E−10	–	–	–	–
RPRC008894	Long Neuropeptide F	–	–	XP_562632	XP_001123033	NP_001127708|	–	–	–	–	–	–	–
KF958188	Myoinhibitory peptide	NP_001280529	NP_001284893	XP_314133	–	NM_001114874	–	–	–	–	–	–	–
GL56309[a]	Myosuppressin	XM_008201934	BT100234	BK001474	EU785050	NM_001134238	–	–	–	–	–	–	–
RPRC001687	Natalisin	NM_001293582	AAA28722	XM_312088	XM_395081	NM_001134276	3E−13	4.00E−44	4.00E−07	–	–	–	–
AFO73269	PBAN	XP_008198650	NP_001014620	AAX84797	DAA05842	NP_001036913	–	–	–	4.00E−06	∞ (0.05)	∞ (0.01)	∞ (0.02)
AFO73270	PBAN	XP_968729	NP_001014620	AAX84797	DAA05842	AEX15646	5.00E−30	8.00E−08	6.00E−52	1.00E−06	∞ (0.05)	∞ (0.01)	∞ (0.02)
AFO73271	PBAN	XP_008198650	AAR30188	XP_003436383	DAA05842	AEX15643	–	–	7.00E−12	2.00E−06	∞ (0.05)	∞ (0.01)	∞ (0.02)
RPRC008528	PBAN	NP_001280528	AAL39284	XP_315694	XP_396491	XP_004929976	–	–	4.00E−06	–	–	–	–
RPRC001000	RFa peptides	XP_008193428	–	–	XP_006563145	–	4E−31	–	–	–	–	–	–
RPRC002266 + RPRC002268 + RPRC002269	Short Neuropeptide F	XP_966794	NP_524176	XP_320180	XP_001123033	NP_001127708	2.00E−48	1.00E−14	4.00E−08	–	–	–	–
RPRC000835	SIFa	XP_970225	NP_732614	DAA35193	XP_396660	XP_00492676	–	–	–	–	–	–	–
GL545664_GL552047_GL562893	Sulfakinins	XM_008193702	BT030304	XM_309215	XM_006562369	NM_001134272	3.00E−08	4E−11	7.00E−11	–	–	–	–
RPRC012816	Sulfakinins	KC161573	AY271617	XM_001237202	XM_006562369	NM_001134272	8.00E−09	–	9.00E−12	–	–	–	–
RPRC003160	Tachykinin	XM_965009	CAA44595	XM_003435839	XM_395081	NM_001134250	4.00E−08	6E−140	1.00E−06	4.00E−07	∞ (0.005)	∞ (0.03)	∞ (0.03)
RPRC014721	Orphan	EFA05746	NP_001096966	–	–	–	–	1E−15	2.00E−16	3.00E−07	∞ (0.04)	∞ (0.06)	∞ (0.11)
RPRC015456	Orphan	EFA10679	NP_650754	XP_565734	XP_001122075	–	–	7E−10	1.00E−10	–	–	–	–

(*continued on next page*)

Table 2 (*continued*)

R. prolixus receptor number	Probable endogenous ligand	*Tribolium castaneum*	*Drosophila melanogaster*	*Anopheles gambiae*	*Apis mellifera*	*Bombix mori*	*Triatoma dimidiata*		*Triatoma infestans*		*Triatoma pallidipennis*		*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
							E-value		*E*-value		*E*-value		*E*-value			
RPRC004128	Orphan	XP_974272	AAF45710	XP_321623	XP_006562927	XP_004928434	–		–		–		–	–	–	–
RPRC008364	Orphan	XP_967232	AAP69822	XP_1688310	–	–	1.00E−29		–		–		2.00E−18	0 (0.74)	0.37 (0.82)	0.37 (0.49)
RPRC011268	Orphan	XP_974272	NP_569970	–	XP_392683	XP_004928327	7.00E−08		2.00E−06		7E−13		–	–	–	–
RPRC004565	Orphan	EEZ99285	NP_001261169	XP_564170	XP_006571770	XP_004932670	–		–		–		–	–	–	–
RPRC004705	Orphan	–	–	XP_321622	–	–	1.00E−34		3.00E−14		5.00E−53		6E−33	0.94 (1)	1.02 (0.99)	1.64 (0.73)

[a] Super contig number is given for transcripts absents from the automatic prediction set. The three last columns indicate fold change and (p) in pairwise comparisons between unfed (basal) and different time points PBM in *R. prolixus* CNS transcriptomes.

Peptide (CCAP) (RPRC001248, RPRC000969 + RPRC012063), Corazonin (CZ) (RPRC000523), ETH (RPRC008652), FaLP (RPRC001551), FaLP/Proctolin (RPRC015267), GPA2/GPB5 (RPRC01428), Kinin (RPRC00494), Long Neuropeptide F (LNF) (RPRC008894, GL563029), Myoinhibiting peptide (KF958188), Myosuppressin (MS) (GL56309) (Leander et al., submitted for publication), Natalisin (RPRC001687) pyroglutamylated RFa peptide (RPRC001000), sNPF (RPRC002266 + RPRC002268 + RPRC002269), SIFa (RPRC000835), PBAN (AFO73269, AFO73270, AFO73271, RPRC008528), Sulfakinins (SK) (RPRC012816, RPRC003273), Tachykinin (TK) (RPRC003160) and orphan receptors or receptors with an unknown ligand (RPRC014721, RPRC015456, RPRC004128, RPRC011268, PRC008570, RPRC000848, RPRC000203). According to our analysis compared with insect orthologous, RPRC001000 is similar in sequence and phylogenetic proximity to pyroglutamylated RFamide receptors (Fig. 2). However, Tanaka et al. (2014) proposed that RPRC001000 would be a thyrotropin releasing hormone receptor (THRH), closely related to FaLPR and ETHR (Tanaka et al., 2014), because it clusters with crustaceans and arachnids THRH.

Orthologous of trissin, RYamide and Inotosin receptors were not found in *R. prolixus*, which fits well with the absence of genes coding for their ligands.

Some ESTs from Triatoma spp. could be clearly assigned to groups according to their putative ligand (Table 2, Fig. 2, Supplementary information 1): TDIM_H9TUR5Q02GDFPZ is a FaLP/proctolin GPCR, TINF_IAZY42G02H0ZZ5 is a CZ GPCR, TDIM_isotig12092 is a CCAP GPCR, TDIM_IAZY42G02GVFWD, TINF_IAZY42G02H4QHT and TPAL H9TUR5Q02IYTLQ are AT GPCR, TINF_isotig37139 is an Allatostatin A GPCR, TDIM_H9TUR5Q02INV8X is an SNF GPCR, TINF_IAZY42G02H6C37 is closely related to Tachykinin GPCRs and TPAL H9TUR5Q01BIIZA is a PBAN GPCR.

3.1.2.2. Family B GPCRs. Family B GPCRs includes calcitonin-like (CT) and CRF-like diuretic hormones (DH) and pigment dispersing factor (PDF). Three CT-DH GPCRs were previously described in *R. prolixus* (Zandawala et al., 2013). Here we report for the first time two new CT-like DH receptors (RPRC004735 and a sequence manually predicted in supercontig GL561032, see Supplementary information 1), four CRF-DH receptors (RPRC015285, RPRC000578 and two sequences manually reconstructed in supercontigs GL562334 and GL563066) and 1 PDF receptor (RPRC009680) (Table 3, Fig. 3). No clear ligand could be assigned for RPRC011083, the only family-B receptor detected in *R. prolixus* CNS, although sequence similarities suggest it could be a Parathyroid hormone receptor (PTHR). PTHRs have been identified in several insects, including the hemiptera *Nilaparvata lugens* (Tanaka et al., 2014). The orthologous of BmNM_001134262 and Tc968807 pair was not found in the present analysis in *R. prolixus*. It is interesting to note that several family-B DH GPCRs seem to be duplicated in *R. prolixus* (Fig. 3). This could reflect the necessity of an effective regulation of water balance in blood sucking bugs that must deal with huge volumes during and after their meal.

We found EST evidence for family-B receptors in *T. dimidata* (7 ESTs), *T. infestans* (2 ESTs) and *T. pallidipennis* (6 ESTs). Sequence similarity and phylogenetic analysis suggest that TPAL_H9TUR5Q01C6ORS is an orthologous of RPRC011083; TINF_IAZY42G01CE3S1, TDIM_H9TUR5Q02I463C, TPAL_H9TUR5Q02GKO1O are CRF-DH receptors and TDIM_isotig15491 is a CT-DH receptor (Table 3, Fig. 3).

3.1.3. Opsins

Color vision in insects is based on the expression of different opsins in photoreceptive cells. Opsins are members of the family A

Table 3
Family-B G protein coupled receptors for neuropeptides identified in *Rhodnius prolixus* genomic information, CNS transcriptome and transcriptomes of Triatoma spp.

R. prolixus receptor number	Probable endogenous ligand	*Tribolium castaneum*	*Drosophila melanogaster*	*Anopheles gambiae*	*Apis mellifera*	*Bombix mori*	*Triatoma dimidiata*	*Triatoma infestans*	*Triatoma pallidipennis*	*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
							E-value	*E*-value	*E*-value	*E*-value			
AHB86317	Calcitonin-like diuretic hormone	XP_008190378	NP_725278	XP_318856	XP_006564466	NP_001127732	1.00E−06	–	–	–	–	–	–
AHB86318	Calcitonin-like diuretic hormone	XP_008190378	NP_725278	XP_318856	XP_006564466	NP_001127735	1.00E−06	–	–	–	–	–	–
AHB86571	Calcitonin-like diuretic hormone	XM_008195698	NP_725278	XM_318856	XP_006564469	NP_001127735	2E−38	–	8E−10	–	–	–	–
GL561032[a]	Calcitonin-like diuretic hormone	XP_008190378	NP_725278	XM_318856	XP_006564469	NP_001127735	–	–	–	–	–	–	–
RPRC004735	Calcitonin-like diuretic hormone	XM_008195698	NM_165979	XM_321982	XM_006564406	NM_001134263	–	–	–	–	–	–	–
GL562334[a]	CRF-like diuretic hormone	XP_008192711	NP_725175	XP_315468	XP_397268	XP_004933474	6E−34	5E−29	3E−56	–	–	–	–
GL563066[a]	CRF-like diuretic hormone	XP_008192711	NP_610960	XP_315468	XP_397268	XP_004933474	2.00E−09	–	2.00E−24	–	–	–	–
RPRC015285	CRF-like diuretic hormone	XP_008192711	NP_610960	XP_315468	XP_397268	XP_004933474	2.00E−09	–	2.00E−24	–	–	–	–
RPRC000578	CRF-like diuretic hormone	XP_008192711	NP_610960	XP_315468	XP_397268	XP_004933474	6.00E−34	5.00E−29	3.00E−56	–	–	–	–
RPRC009680	Pigment Dispersing Factor	XP_008193265	AHN59297	XP_313426	XP_006559533	NP_001127733	–	–	–	–	–	–	–
RPRC011083	Orphan	XP_969953	–	–	XP_006566899	–	5E−16	–	–	3.00E−23	0 (0.95)	0 (0.71)	0.21(0.61)

[a] Super contig numbers are given for transcripts absents from the automatic prediction set. The three last columns indicate fold change and (p) in pairwise comparisons between unfed (basal) and different time points PBM in *R. prolixus* CNS transcriptomes.

Table 4
Opsins G protein coupled receptors identified in *Rhodnius prolixus* genomic information, CNS transcriptome and transcriptomes of Triatoma spp.

R. prolixus receptor number	Probable endogenous ligand	*Tribolium castaneum*	*Drosophila melanogaster*	*Anopheles gambiae*	*Apis mellifera*	*Bombix mori*	*Triatoma dimidiata*	*Triatoma infestans*	*Triatoma pallidipennis*	*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
							E-value	*E*-value	*E*-value	*E*-value			
RPRC010623	–	XP_973147	–	XP_003435763	XP_397397	XP_004928488	–	–	–	–	–	–	–
RPRC002621	–	XP_970344	NP_524411	XP_001688790	XP_392791	XP_004931633	2.00E−44	–	6.00E−16	1.00E−121	0.95 (1)	1.02 (0.99)	1.64 (0.73)
RPRC001048 + RPRC001049	–	XP_001816446	–	XP_312502	NP_001035057	XP_004928383	–	–	–	2.00E−08	0.95 (1)	1.02 (0.99)	1.64 (0.73)
RPRC015283	–	XP_970344.1	NP_524035	XP_308329	XP_392791	XP_004925237	5.00E−07	5.00E−08	2.00E−12	8.00E−25	0.67 (0.90)	0.99 (1)	1.73 (0.71)

The three last columns indicate fold change and (p) in pairwise comparisons between unfed (basal) and different time points PBM in *R. prolixus* CNS transcriptomes.

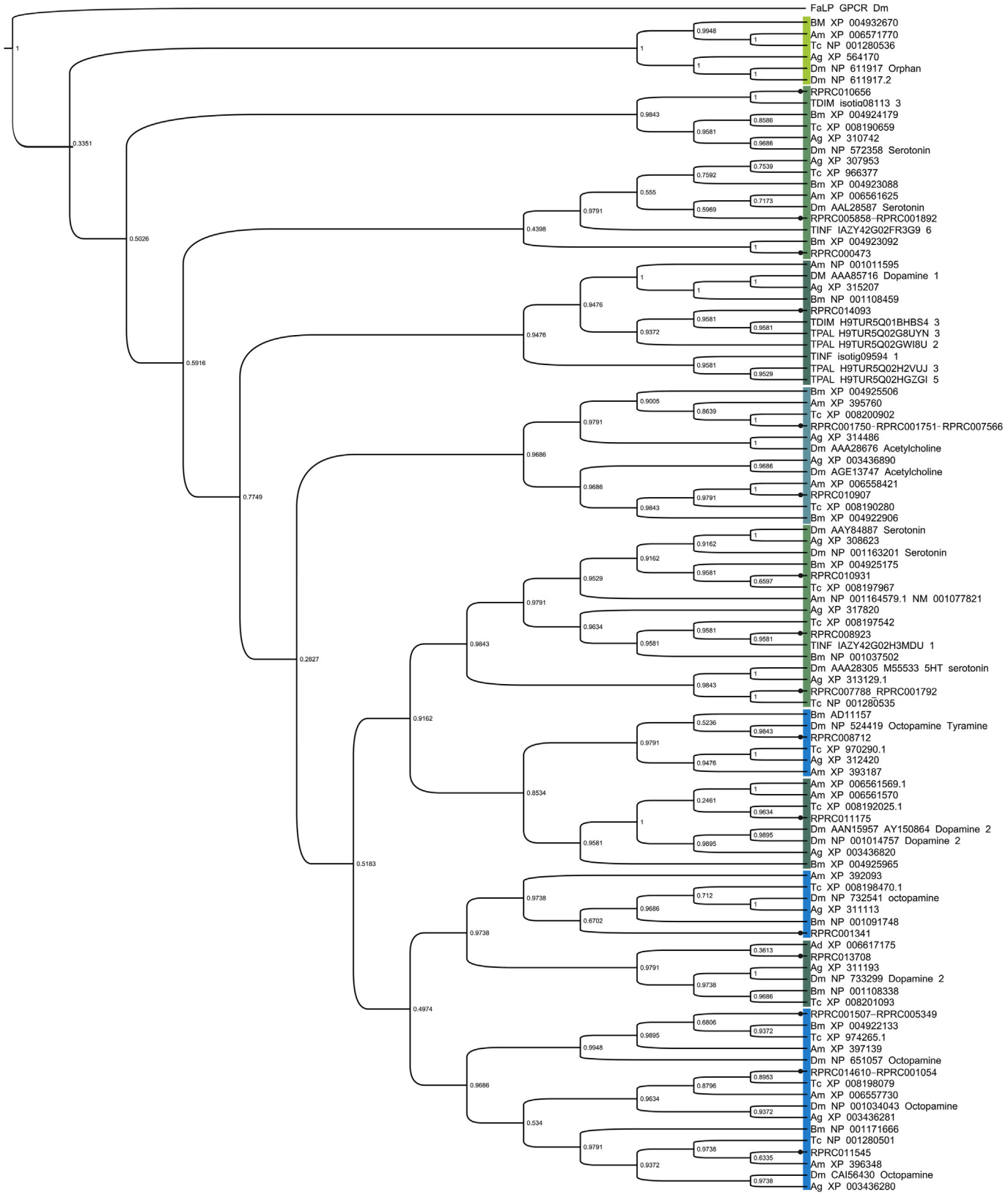

Fig. 1. Phylogenetic tree analysis of biogenic amine GPCRs from *D. melanogaster* (Dm), *A. mellifera* (Am), *A. gambiae* (Ag), *T. castaneum* (Tc), *B. mori* (Bm), *R. prolixus* (RPRC), *T. dimidiata* (TDIM), *T. infestans* (TINF) and *T. pallidipennis* (TPAL). The tree is rooted by *D. melanogaster* metabotropic glutamate receptor (AAF59402). Assigned ligands for *D. melanogaster* are highlighted with different colors. The number at each node indicates the posterior probabilities. *R. prolixus* sequences are indicated with a black circle.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

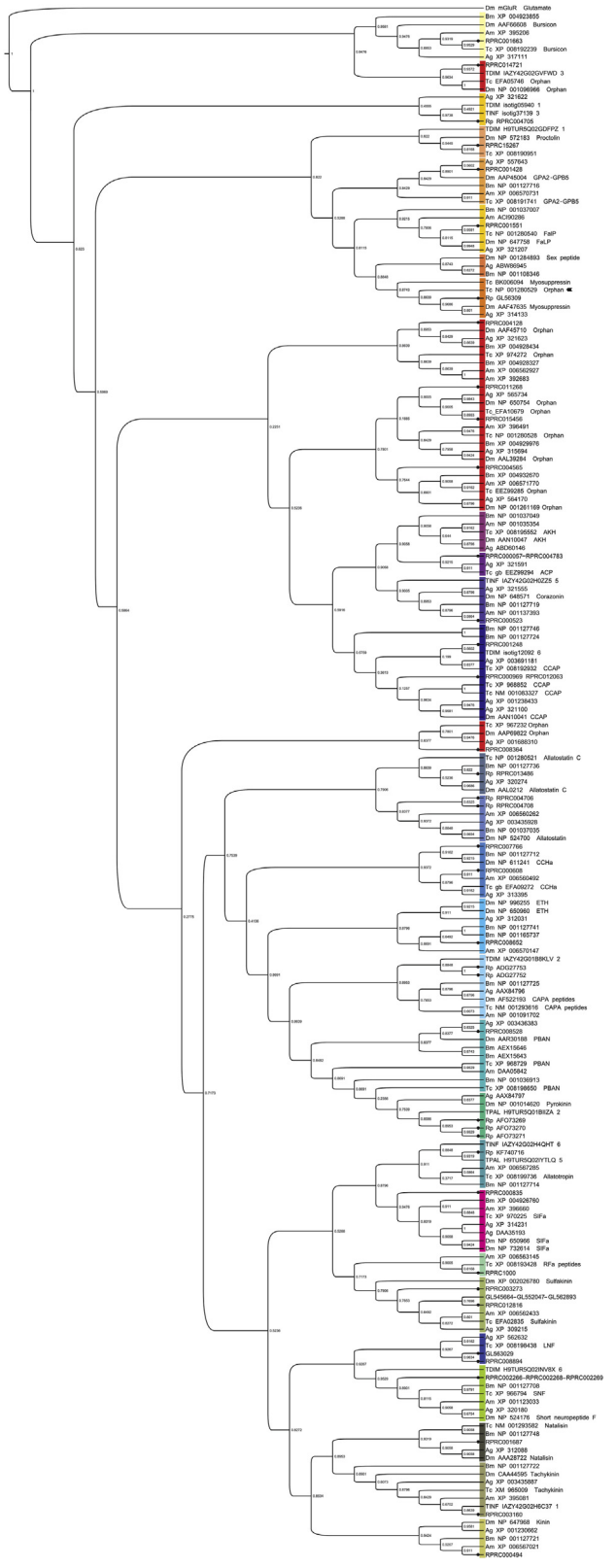

Fig. 2. Phylogenetic tree analysis of family-A neuropeptide GPCRs from *D. melanogaster* (Dm), *A.mellifera* (Am), *A. gambiae* (Ag), *T. castaneum* (Tc), *B. mori* (Bm), *R. prolixus* (RPRC) and *T. dimidiata* (Td). The tree is rooted by *D. melanogaster* metabotropic glutamate receptor (AAF59402). Assigned ligands for *D. melanogaster* and *T. castaneum* are indicated with different colors. The number at each node indicates the posterior probabilities. *R. prolixus* sequences are indicated with a black circle.

GPCRs and are found coupled to light-sensitive chromophores in animal photoreceptors. Three groups of opsins have been reported in *D. melanogaster*: one related to long-wavelength vision (including Rh1, Rh2 and Rh6), another group related to short-wavelength vision (Rh3, Rh4 and Rh5), and a third group including only Rh7 (Brody and Cravchik, 2000; Salcedo et al., 1999). On the other hand, a fourth group of invertebrate opsins named pteropsins is missing from the genome of drosophilid flies (Velarde et al., 2005).

Using phylogenetic analysis and sequence similarity, 4 putative opsins from *R. prolixus* were identified for the first time. From these, 3 have EST evidence in *R. prolixus* CNS, 2 in *T. dimidiata*, 1 in *T. infestans* and 2 in *T. pallidipennis* (Table 4, Fig. 4). The phylogenetic tree (Fig. 4) suggests that RPRC015283 is related to the first class, RPRC002621 and TDIM_AZY42601CQ0F5 are related to the second class, RPRC015283 belongs to the third class, and RPRC010623 is close to pteropsins. Together, the results provide evidence of genes from the four families of invertebrate opsins in *R. prolixus* genome.

3.1.4. Neurotranscriptome of *R. prolixus* nervous system

cDNA libraries derived from *R. prolixus* CNS at different times post feeding were sequenced on the HiSeq 2000 platform; 74.7×10^6 transcriptomic reads were generated. Illumina reads were assembled in 97,071 locus with different splicing variants (55,970 unique transcripts) totaling 29,863,582 bp. The locus length varied from 100 to 5734 bp (Supplementary information 2), in which the mean contig length was 308 bp and the N50 was 231 bp (Table 5). To estimate the coverage and representation degree of the neurotranscriptome, we analyzed three parameters: 1) the core eukaryotic genes (CEG), 2) transcript mapping to the *T. dimidiata* mitochondrial genome and 3) alignment with *R. prolixus* genome.

The CEG coverage was calculated for *R. prolixus* CNS transcriptome searching for the 458 core eukaryotic protein models (Parra et al., 2007). We identified 447 CEGs in our neurotranscriptome dataset, corresponding to 97.6% of coverage.

The mitochondrial *R. prolixus* genome on the CNS neurotranscriptome showed high degree of conservation compared to the mitochondrial genome of *T. dimidiata*; 11,828 bp aligned around 70% of similarity. This result shows that the most conserved genes in eukaryotes are present and highly conserved in the CNS transcriptome of *R. prolixus*, and that the mitochondrial genome is nearly complete.

Since *R. prolixus* genome is completely sequenced, we performed an alignment of *R. prolixus* neurotranscriptome and *R. prolixus* genomic sequence to obtain further parameters of transcript assembly accuracy and integrity. The current assembled sequences of *R. prolixus* genome cover 95% of the genome, ~702 MB, with 15,441 putative protein encoding genes. Using the CNS transcripts to map the genome sequence, 47,095 sequences (~50% of the assembled transcripts) aligned with *R. prolixus* genome. These aligned sequences comprise 14,035,192 bp, approximately 2% of *R. prolixus* genomic sequence.

Our experimental design implied extracting RNA at different physiological time points PBM, in order to get the wider representation possible of the CNS mRNAs, as it has been previously done with the gut transcriptome (Ribeiro et al., 2014). This enabled us to make a differential introductory expression analysis of genes of interest. Here we present the results obtained by making pairwise comparisons between starved conditions (basal), 1 h, 4 h and 24 h PBM of neuropeptide precursors and GPCRs genes (Tables 1–4 and 6). Even though further experiments specifically oriented to extract conclusions about differential gene expression after blood meal should be performed, the results suggest there is a regulation of some GPCRs after a blood meal. Indeed, previous results on mass

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

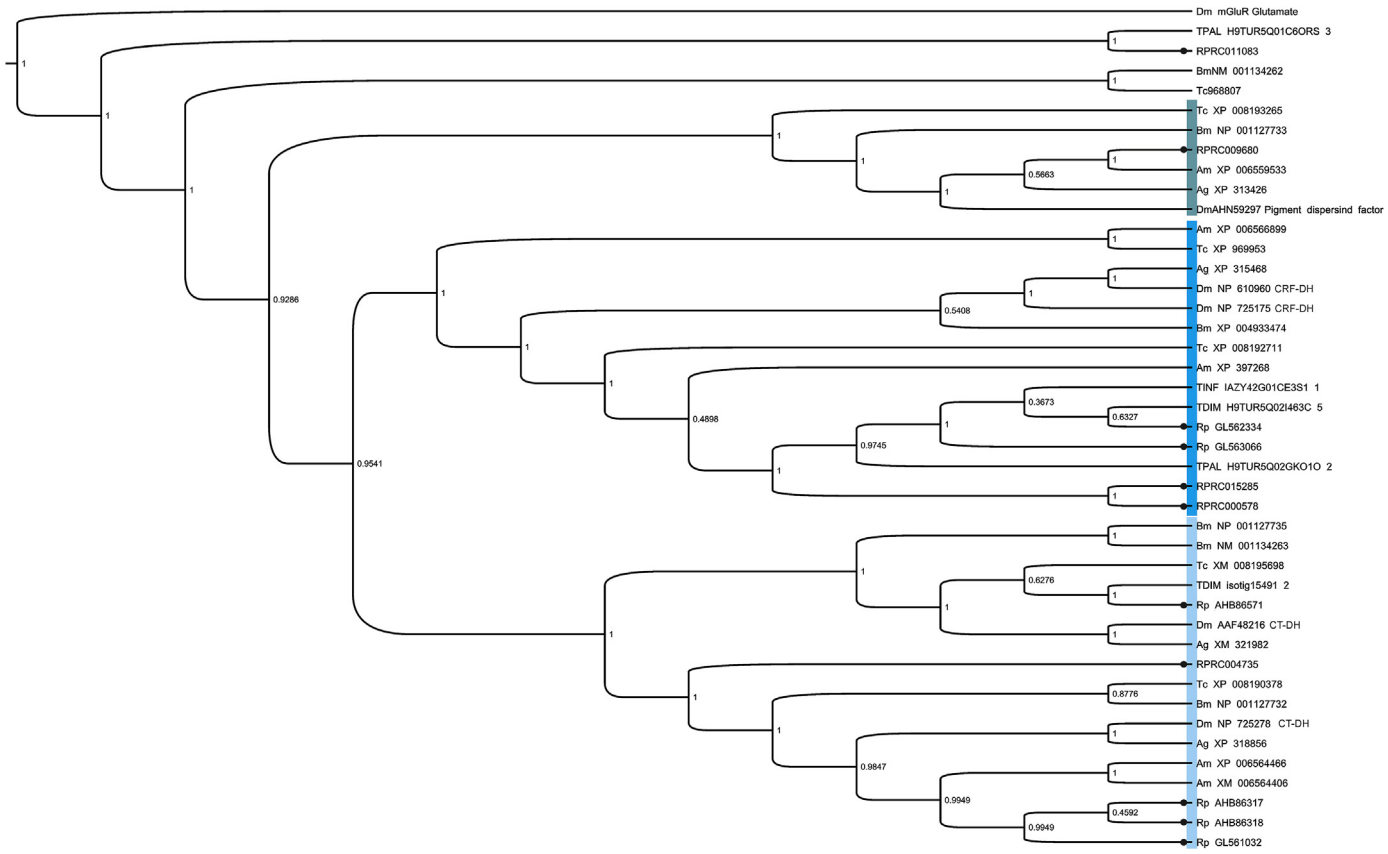

Fig. 3. Phylogenetic tree analysis of family-B neuropeptide GPCRs from *D. melanogaster* (Dm), *A. mellifera* (Am), *A. gambiae* (Ag), *T. castaneum* (Tc), *B. mori* (Bm), *R. prolixus* (RPRC) and *T. dimidiata* (Td). The tree is rooted by *D. melanogaster* metabotropic glutamate receptor (AAF59402). Assigned ligands for *D. melanogaster* are indicated with different colors. The number at each node indicates the posterior probabilities. *R. prolixus* sequences are indicated with a black circle.

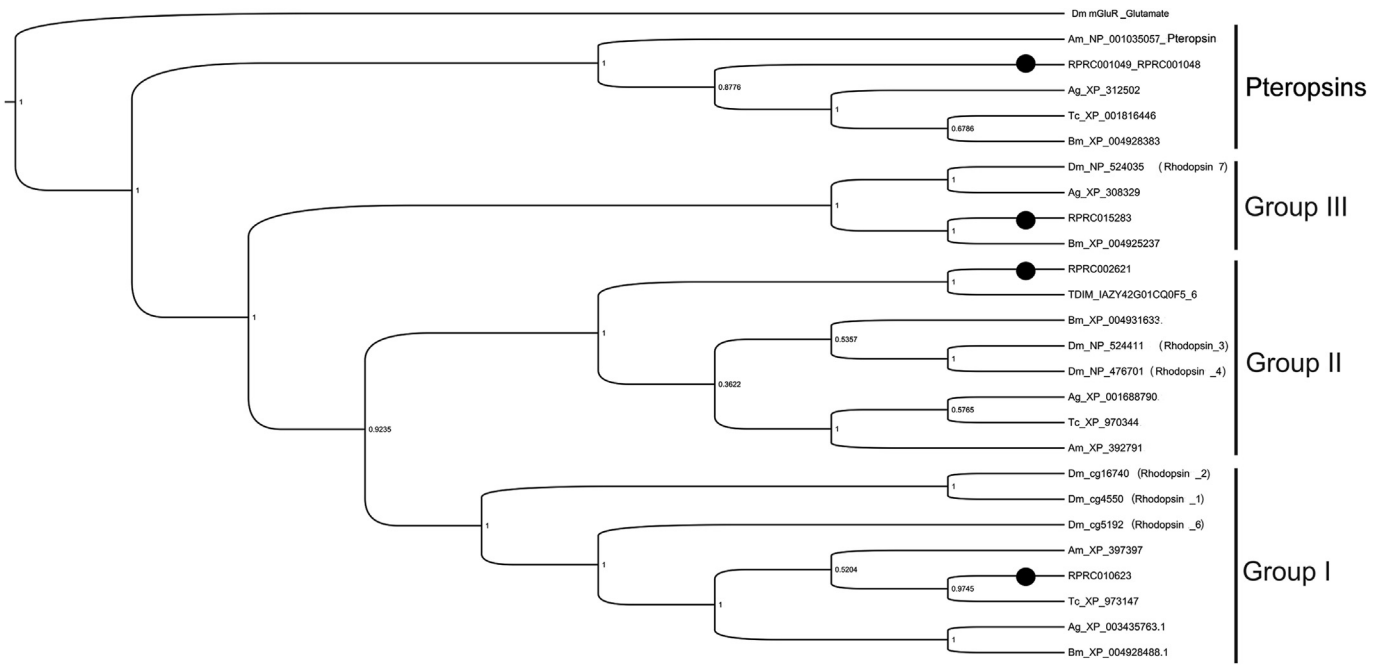

Fig. 4. Phylogenetic tree analysis of opsins GPCRs from *D. melanogaster* (Dm), *A. mellifera* (Am), *A. gambiae* (Ag), *T. castaneum* (Tc), *B. mori* (Bm), *R. prolixus* (RPRC), *T. dimidiata* (Td), *T. infestans* (Ti) and *T. pallidipennis* (Tp). The tree is rooted by *D. melanogaster* metabotropic glutamate receptor (AAF59402). Assigned ligands for *D. melanogaster* or *A. mellifera* are shown in parentheses. The number at each node indicates the posterior probabilities. *R. prolixus* sequences are indicated with a black circle.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

Table 5
Rhodnius prolixus CNS transcriptome libraries sequencing and assembly metrics.

	Illumina reads
A. Reads by samples	
Basal condition	15.9×10^6
1 h PBM	12.6×10^6
4 h PBM	18.6×10^6
24 h PBM	27.5×10^6
Total	74.7×10^6
B. Transcript assembly summary	
Transcripts assembled	55,970
Locus	97,071
Mean contig length (bp)	308
N50 (bp)	231
Largest transcript (bp)	5734
Total bases (Mbp)	29
CEG	97.60%
Mitocondrial genome cov. (%)	69.70%
Transcript mapped to *R. prolixus*	47095

spectrometry-based differential neuropeptide expression in *R. prolixus* CNS showed that differences are subtle, as one might expect in physiological processes (Sterkel et al., 2011). Furthermore, neuropeptide activity can also be regulated by post translational modifications.

3.1.5. Comparative analysis

To find putative orthologous, we used BLASTX searches among *R. prolixus* neurotranscriptome, the predicted proteomes of different arthropods species, and the brain transcriptome of *S. gregaria* (Badisco et al., 2011). The analysis revealed high putative orthologue conservation; we found 21,540 hits (4197 unique hits, ignoring different splicing variants that matched with the same subject protein) in *R. prolixus* database, 20,482 (3239 unique hits) in *A. pisum*, 16,065 (4094 unique hits) in *S. gragaria*, 11,402 (2239 unique hits) in *A. albimanus*, 10,844 (2481 unique hits) in *A. aegipty*, 9899 (2420 unique hits) in *A. gambiae*, 12,631 (2407 unique hits) in

Table 6
Neuropeptide precursors identified in transcriptomes.

	Query ID	*Triatoma dimidiata*	*Triatoma infestans*	*Triatoma pallidipennis*	*Rhodnius prolixus* CNS	1 h PBM vs Basal	4 h PBM vs Basal	24 h PBM vs Basal
		E-value	*E*-value	*E*-value	*E*-value			
Neuropeptides								
ACP	GL553060[a]	–	–	–	2.00E−14	0.32 (0.60)	0.34 (0.52)	1.20 (0.91)
Adipokinetic hormone	GU062794.1	7.00E−16	–	–	1.00E−36	0.38 (0.58)	0.51 (0.66)	0.73 (0.82)
Allatostatin A	GQ856315	–	–	–	5.00E−49	0.89 (1)	0.73 (0.87)	1.28 (0.87)
Allatostatin C	GL562620[a]	3.00E−48	–	–	1.00E−54	0 (1)	inf (0.7)	0 (1)
Allatostatin CC	RPRC000300	8.00E−10	4.00E−10	5.00E−10	2.00E−09	0 (1)	0 (1)	0 (1)
Allatotropin	GQ162783	1.00E−40	–	–	5.00E−49	0 (1)	0 (1)	0 (0.9)
Bursicon alpha	RPRC000797	–	–	9.00E−47	–	–	–	–
Bursicon beta	GL563191[a]	–	6.00E−24	–	7.00E−49	0.66 (1)	0.65 (1)	0.046 (0.66)
Calcitonin-like diuretic hormone A	GQ856316	–	1.00E−33	4.00E−37	–	–	–	–
Calcitonin-like diuretic hormone B	GQ856317	–	2.00E−24	1.00E−32	–	–	–	–
CCH-amide peptide	GL562918[a]	1.00E−21	1.00E−21	–	3.00E−36	0 (1)	0 (0.60)	0.32 (0.77)
CNM-amide peptide	RPRC010893	–	–	–	–	–	–	–
Corazonin	GL563001[a]	–	–	–	7.00E−45	0.42 (0.73)	0.70 (0.87)	1.92 (0.67)
CRF like Diuretic Hormone	HM153808	8.00E−79	–	–	–	–	–	–
Crustacean Cardioactive peptide	GQ888668	–	–	–	6.00E−63	0 (1)	0 (1)	0.75 (0.85)
Eclosion hormone	RPRC014242	–	–	–	–	–	–	–
Elevenin-1	RPRC003083	–	–	–	–	–	–	–
Elevenin-2	RPRC003084	–	–	–	–	–	–	–
ETH	RPRC014486	7.00E−19	3.00E−22	6.00E−19	–	–	–	–
FaLP	RPRC014988	–	–	–	5.00E−15	1.3 (0.95)	0.82 (1)	0.65 (0.82)
Insect kinin	BK007870	–	–	–	–	–	–	–
Insulin-like peptide 1	RPRC009717	–	–	–	5.00E−09	2.7 (0.5)	1.79 (0.68)	2.55 (05)
Insulin like peptide 2	GU230850	–	1.00E−08	–	1.00E−35	0 (1)	0 (1)	0 (1)
Insulin like peptide 3	GL562992[a]	–	–	–	–	–	–	–
Insulin like peptide 6	GL563143[a]	–	–	–	6.00E−07	0 (1)	0 (1)	0 (1)
Ion Transport peptide A	GQ253921	–	–	–	–	–	–	–
Ion Transport peptide B	GU207866	–	–	–	–	–	–	–
ITG-like	GL562724[a]	–	1.00E−21	7.00E−21	2.00E−14	0.76 (0.97)	0.80 (0.93)	0.62 (0.75)
Long Neuropeptide F	RPRC008107-RA	–	–	–	1.00E−59	0.51 (0.71)	0.45 (0.59)	0.63 (0.75)
Myoinhibiting peptide	GL562915[a]	5.00E−42	–	–	–	–	–	–
Myosuppressin	GQ344501	–	1.00E−10	–	–	–	–	–
Natalisin	RPTMP02028	5.00E−50	–	4.00E−54	1.00E−55	2 (0.64)	2.72 (0.50)	2.37(0.55)
Neuroparsin A	GU207864	3.00E−51	3.00E−57	1.00E−57	4.00E−61	0.85 (1)	0.78 (0.90)	1.39 (0.83)
Neuropeptide-like precursor 1	GU207865	1.00E−18	2.00E−23	2.00E−06	0.00E+00	0 (1)	0 (0.88)	0.89 (0.95)
NVP-like	RPRC000343	9.00E−82	4.00E−66	–	–	–	–	–
Orcokinin A	FJ167860	2.00E−80	7.00E−82	2.00E−76	2.00E−28	0 (1)	0 (1)	0 (0.78)
Orcokinin B	JF761320	2.00E−20	1.00E−20	1.00E−42	1.00E−24	0 (1)	0 (1)	1.05 (0.99)
PBAN	GU230851	–	–	–	3.00E−24	0.42 (0.70)	0 (1)	1.05 (0.99)
Pigment dispersing factor	GL563064[a]	–	–	–	6.00E−17	1 (0.96)	0.93 (0.97)	1.13 (0.93)
Proctolin	JN543225	–	–	5.00E−11	1.00E−41	0.62 (0.81)	0.47 (0.62)	1.20 (0.90)
Short Neuropeptide F	GQ452380	6.00E−47	3.00E−30	4.00E−46	3.00E−50	0 (1)	0 (0.84)	0 (0.31)
SIF-amide	GQ253922	–	–	–	3.00E−38	–	–	–
Sulphakinins	GQ162784	–	–	–	2.00E−43	–	–	–
Tachykinins	GQ162785	–	–	–	1.00E−59	–	–	–

The three last columns indicate fold change and (p) in pairwise comparisons between unfed (basal) and different time points PBM in *R. prolixus* CNS transcriptomes.
[a] Supercontig number in *R. prolixus* genome sequence.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

Table 7
BLASTX results comparing the *Rhodnius prolixus* CNS transcriptome with other orthropods proteomes.

	Number of hits	Uniq hits	Locus wit hit (%)	Average identity (%)	Reference proteome size
R. prolixus	21,540	4198	22.2	77.7	15,441
A. albimanus	11,402	2239	11.7	54.3	11,994
A. aegypti	10,844	2481	11.2	58.9	17,408
A. gambiae	9899	2420	10.2	59.8	14,667
C. quinquefasciatus	12,631	2407	13	57.8	19,018
A. pisum	20,482	3239	21.1	53.9	33,291
D. melanogaster	16,767	2896	17.3	53.8	79,273
I. scapularis	7520	2016	7.7	55.4	20,486
P. humanus	10,047	2332	10.3	60.4	10,774
L. longipalpis	11,911	1995	12.3	54.1	10,110
P. popatasi	8265	2146	8.5	54.7	11,164
S. gregaria	16,065	4094	16.5	56	76,254

Culex quinquefasciatus, 16,767 (2896 unique hits) in *D. melanogaster*, 7520 (2016 unique hits) in *I. scapularis*, 10,047 (2332 unique hits in) *P. humanus*, 11,911 (1995 unique hits) in *L. longipalpis* and 8265 (2146 unique hits) in *P. papatasi* (Table 7). The number of orthologue identification and identity average was higher for *R. prolixus* (22% of the locus match with *R. prolixus* proteome and 77.7% of identity average), but it was also significant in the searches performed on *S. gregaria* and *A. pisum* – matching 16.5 and 21.1% locus, with an average identity of 56 and 53.9% respectively. The observed difference in the orthologue representation seems to correspond to the evolutionary and functional characteristics among these databases and the neurotranscriptome dataset: our dataset was made from an *R. prolixus* CNS library, *A. pisum* is an hemipteran closely related to *R. prolixus*, and *S. gragaria* database is a brain transcriptome.

Thereby, the library metrics showed here – 2% of *R. prolixus* genome coverage 4197 unique BLASTx hits comparing to *R. prolixus*, 97.6% of the CEG and 69.7% of the mitochondrial genome coverage – are satisfactory parameters of completeness and high conservation levels of our CNS transcriptome in eukaryotes, as well as in the genomic information available for *R. prolixus*.

3.1.6. Central nervous system orthologues

In order to find orthologues genes implied in neurosecretory events we performed a bidirectional BLASTX search to find the best reciprocal hit between the *S. gregaria* CNS transcriptome and the neurotranscriptome of *R. prolixus*. The bidirectional comparison revealed 3422 best reciprocal hits, representing 6.2% of the predicted proteome for *R. prolixus* and 4.5% for *S. gregaria*. The low proportion of orthologues can be explained taking into account that both protein datasets are the translate transcriptome (including the six possible reading frames), and not predicted proteomes. On the other hand, many transcripts in the datasets are from untranslated regions, areas with low conservation levels and transcripts with assembly errors. Beyond that, we have found many sequences that could be orthologues in the CNS of *R. prolixus* and *S. gregaria* (Supplementary information 3).

3.1.7. Gene ontology annotation

Using Blast2Go, we assigned Gene Ontology (GO) terms to 2518 transcripts in our CNS transcriptome which were distributed in the three main GO categories: *Biological Process*, *Molecular Function* and *Cellular component* (Fig. 5). In *Biological Process* category, the best represented sub-categories were *Cellular process* (695 sequences), *Metabolic process* (593 sequences) and *Single organism cellular process* (284 sequences) (Fig. 5A). To obtain a more detailed view in these GO groups, we looked at the sequence distribution in a multilevel pie (Fig. 6A). This showed that the biggest sequence groups in the *Biological process* category were *DNA metabolic process* (n = 136), *Signal transduction* (n = 107), *Cellular protein modification process* (n = 99), *Cellular amino acid metabolic process* (n = 85) and *Transmembrane transport* (n = 65). The sequences grouped in *Cellular protein modification process* and *S. transduction* represented approximately 24% of *Cellular process* GO term, being one of the most important components in this group. In the same way, *DNA metabolic process* and *Cellular amino acid metabolic process* represented two of the most important sequences (37%) grouped in the *Metabolic process*. *Transmembrane transport* is the only GO term with high number of sequences in the *Single-organism process* GO level, representing approximately 27% of the sequences on this level.

Molecular function GO term is mainly conformed by three GO sub-categories: *Catalytic activity* (500 sequences), *Binding* (491 sequences) and *Transporter activity* (61 sequences) (Fig. 5B). The multilevel pie showed that most of the sequences grouped in *C. activity* (48%) were classified in *Oxidoreductase activity* (n = 92), *Kinase activity* (n = 69), *Nucleotidyltransferase activity* (n = 45) and *Methiltransferase activity* (n = 35) (Fig. 6B). On the other hand, *Binding* GO sub-category is mainly composed of sequences corresponding to the sub-categories *Ion binding* (n = 360) and *DNA binding* (n = 92), which represent 92% of the sequences classified in *Binding* GO term (Fig. 6B). *Transporter activity* – with lower number of sequences than the categories already mentioned – is the third largest group into *Molecular function* GO term with 61 sequences of which 97% (n = 59) belong to *Transmembrane transporter activity* GO term (Fig. 6B).

Strikingly, many of the largest GO terms identified correspond to gene groups in which some genes are involved in Neuropeptide and GPCRs synthesis pathways or hormone receptors pathways. Thereby, many neuropeptides, peptide hormones and their receptors - such as Tachynins, Miosupresins, Sex peptides, Opsines, Serotonin, Octopamines, Acetylcholine, Proctolin and Dopamines – are included into *Signal transduction GO family* (*Biological process* → *Cellular process* → *Cell communication* → *Signal transduction*), *Transmembrane transporter activity* (*Molecular function* → *Transporter activity* → *Transmembrane transporter activity*), or *Transmembrane transport* (*Single-organism process* → *Single-organism transport* → *Transmembrane transport*). These results show an enrichment in molecular, metabolic and cellular pathways involved in neurophysiology and neurosecretory events.

3.2. Neuropeptide and peptide hormones

We performed database searches in *R. prolixus* nervous system and *T. dimidiata*, *T. infestans* and *T. pallidipennis* normalized transcriptomes using the sequences of *R. prolixus* neuropeptide precursor genes previously described as queries (Ons et al., 2011; Orchard et al., 2011; Sterkel et al., 2012). Orthologous sequences

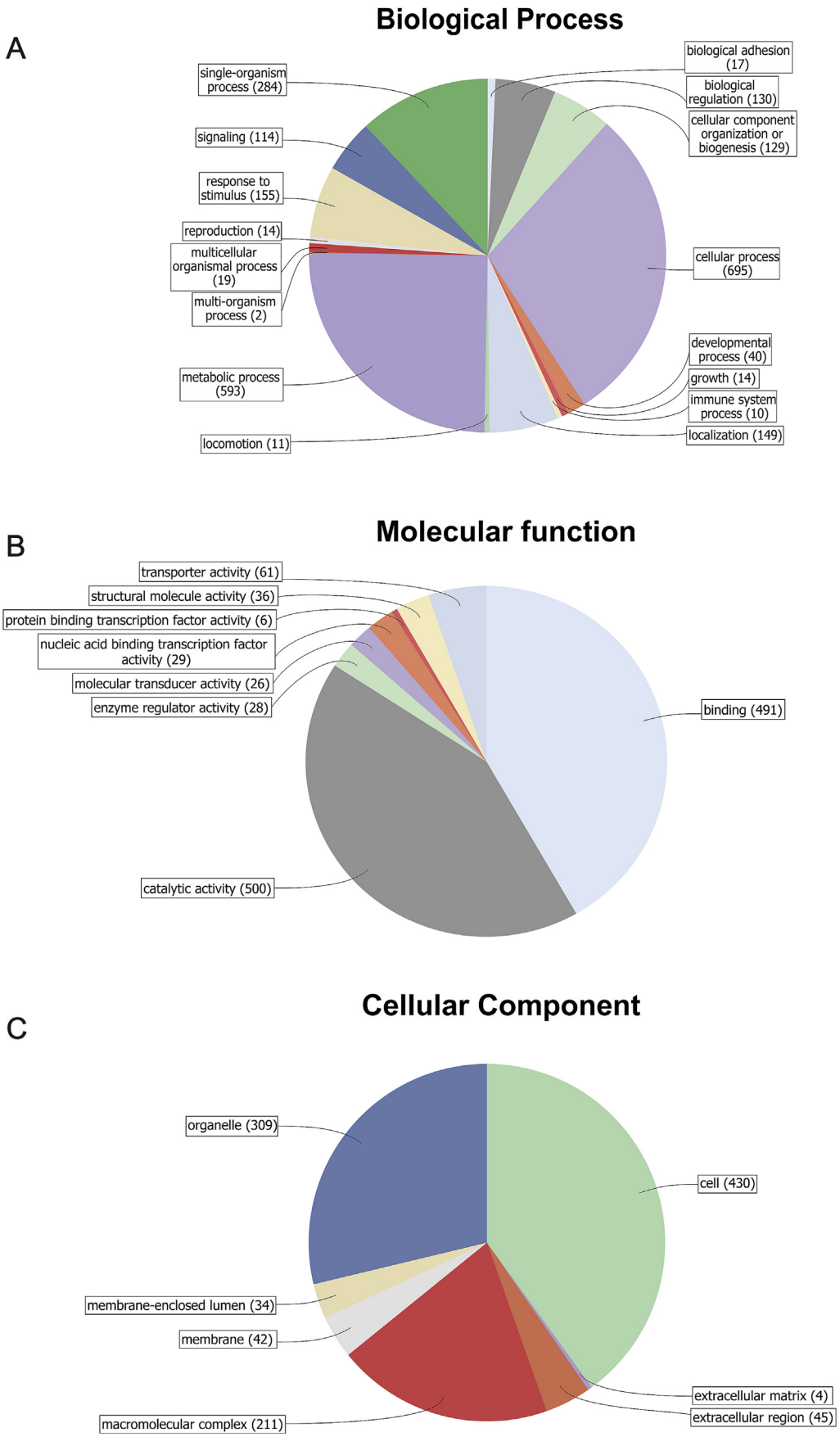

Fig. 5. Gene Ontology (GO) terms assignation. *R. prolixus* CNS transcripts distribution on the three main GO domains: Biological processes (A), Molecular function (B) and Cellular components (C).

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

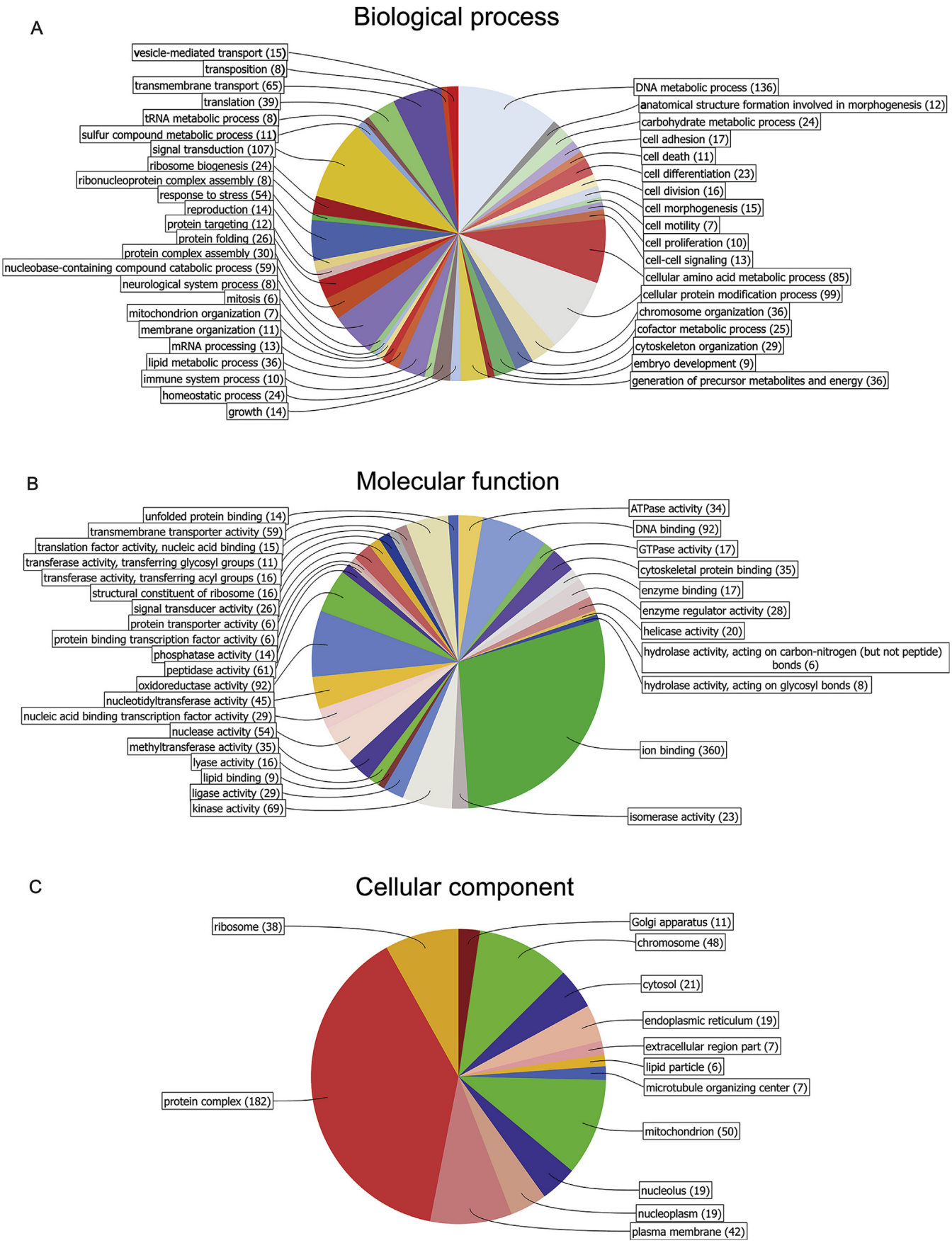

Fig. 6. Sequence distribution by GO terms. Multilevel pies in which the number of sequences on the lowest nodes in the GO analysis of Biological process (A), Molecular function (B) and Cellular component (C) are shown.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003

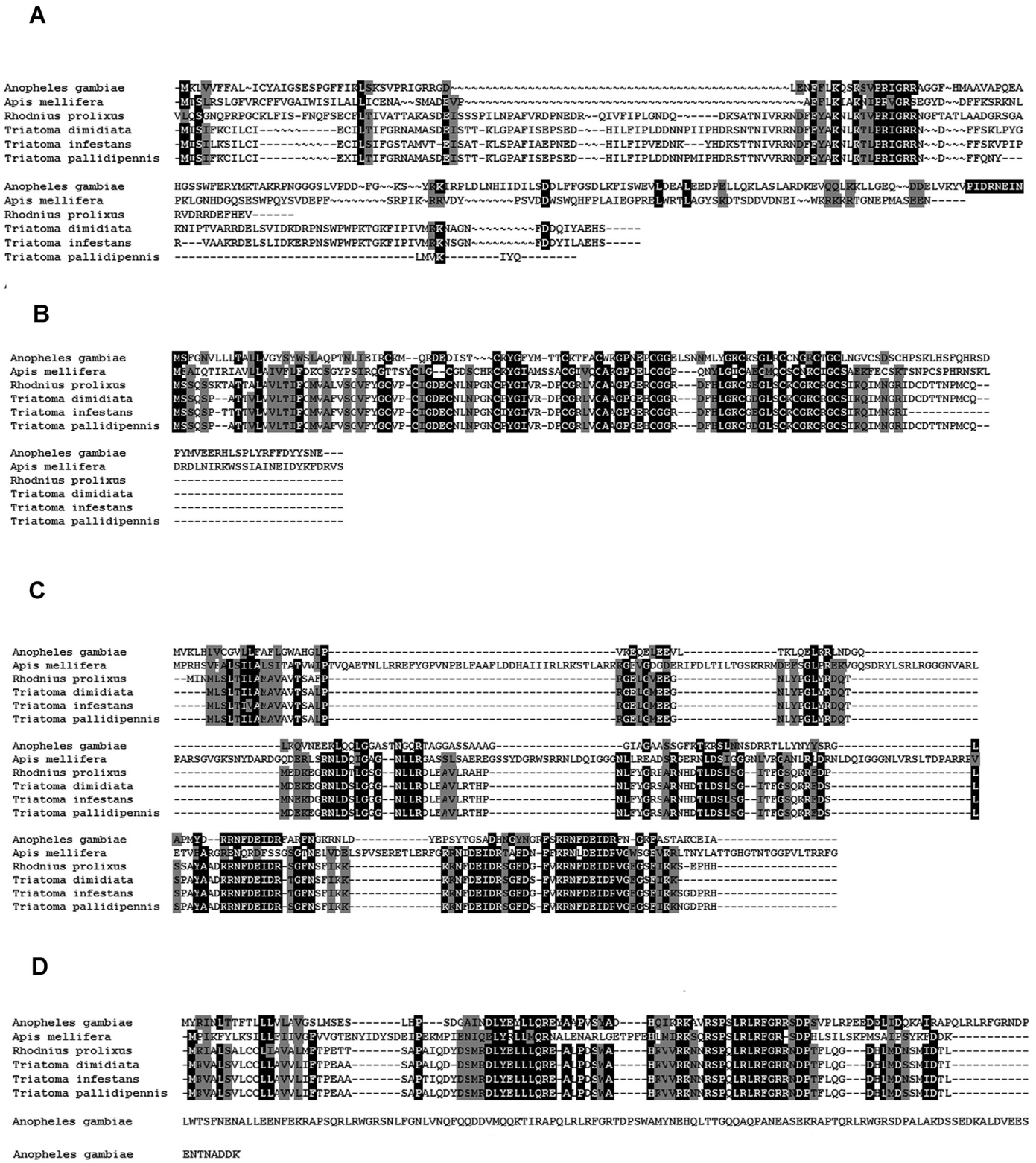

Fig. 7. Alignment of neuropeptide precursor genes from *A. gambiae*, *A. mellifera*, *R. prolixus*, *T. dimidiata*, *T. infestans* and *T. pallidipennis*. Putative proteolitic cleavage sites predicted according to Veenstra (2000) are boxed. Amino acid residues that are identical in all four sequences are highlighted in black; similar amino acid residues in all four sequences are highlighted in gray. The core peptide is underlined in black. Signal peptide is underlined in gray. A) Ecdysis triggering hormone; B) Neuroparsin-A precursor; C) Orcokinin A; D) Short Neuropeptide-F.

from other insects were used for those neuropeptides that had not been identified in *R. prolixus* yet.

We identified 25 of the 39 previously described *R. prolixus* neuropeptide precursor genes (NPGs) in the nervous system transcriptome (Table 6). From the different *Triatoma* spp., we found 15 NPGs in *T. dimidiata*, 14 in *T. infestans* and 12 in *T. pallidipennis* (Table 6). The differences in the efficiency of NPGs detections arose because in the nervous system neuropeptides (*R. prolixus*) they are

more represented than in the normalized transcriptomes (*Triatoma* spp.).

Some particularities detected in *R. prolixus* compared to other insects seem to be characteristic of triatomines. It is the case, for instance, for Myosuppressin (MS); in the highly conserved C-terminal region of *R. prolixus* MSP the active site shows an FMRF-NH2 (−NH2═C terminal amidation) instead of the FLRF-NH2-C terminal domain, which is characteristic of the MSPs described so far (Nässel, 2002). This particularity is also present in the MSP identified in *T. dimidiata*, indicating that it is probably a characteristic of triatomine MSPs (see sequence TDIMH9TUR5Q02IDL4O in Supplementary Information 4). Moreover, we were not able to detect the following neither in *R. prolixus* genome assembly nor in the transcriptomes analyzed: CCHamide 2, inotosin, prothoracicotropic hormone (PTTH), RYamide, Sex peptide (SP), SMYamide nor trissin. PTTH is a neurohormone involved in the control of development in insects, absent in the hemimetabolous genomes and transcriptomes described so far, with the only exception of *N. lugens* (Tanaka et al., 2014). However, the existence and activity of PTTH in *R. prolixus* was suggested by experimental data (Vafopoulou and Steel, 2014; Vafopoulou et al., 2007). We cannot rule out highly diverging sequences that were not detected by our homology search.

Five NPGs were fully or partially detected in *R. prolixus* genome as well as in the three transcriptomes from triatoma spp., probably indicating a high expression: Ecdysis Triggering Hormone (ETH) (Fig. 7A), Neuroparsin-A (NPA) (Fig. 7B), Orcokinin A (OKA) (Fig. 7C), Short Neuropeptide F (sNPF) (Fig. 7D) and Neuropeptide-like precursor 1 (NPLP1). The comparisons of orthologous NPGs among the four triatomine species indicate high degree of identity in the bioactive peptides, and a higher degree of variation in the spacer region (the parts separating the bioactive peptides). This fact was also observed by Wegener and Gorbashov in their comprehensive analysis of the molecular evolution of neuropeptides in the genus Drosophila (Wegener and Gorbashov, 2008), suggesting that spacer regions are not involved in signaling, where single amino acid changes could result in an altered efficacy of the molecule. Less selection pressure seems to affect spacer regions. However, spacer peptides can be important in proper peptide processing and/or modulation of neuropeptide activity, as suggested by experimental evidence (Brezden et al., 1999).

4. Conclusion

Our genomic and transcriptomic analysis of neuropeptide and GPRCs for neurohormones and opsines provides information of genes from *R. prolixus*, *T. dimidiata*, *T. infestans* and *T. pallidipennis*. The results suggest a high degree of conservation of the neuroendocrine system in triatomines at the sequence level.

Basing on comparisons with other well-studied insects, there seems to be no significant differences in the number of GPCRs between holometabolous and hemimetabolous (Hauser et al., 2008, 2006; Li et al., 2013; Tanaka et al., 2014). Moreover, the hormonal systems are conserved between holometabolous and hemimetabolous, with the only exception of trissin, which seems to be lost in all the hemimetabolous species reported to date. However, the physiological role of trissin remains unclear. The data presented here is the first comprehensive analysis of GPCRs and CNS-expressed genes in triatomines. We consider that this data will contribute to boost physiological genomic research in the vectors of Chagas disease.

Acknowledgments

We thank Raúl Stariolo from Centro de Referencia de Vectores, Punilla, Córdoba, Argentina for the help in the providing of insects, and to Verónica Kicinski for the editorial revision. This work was supported by grants from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET Grant PIP 2012-2014 049 to S.O) and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT2010-0135 and PICT2013-1554 to R.R.P. and PICT 2011-194 to S.O). S.O. and R.R.P. are investigators from CONICET. J.P.W. and A.L. are recipients of research fellowships from CONICET. I.S. is recipient of a research fellowship from ANPCyT.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2015.05.003.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., et al., 2000. The genome sequence of Drosophila melanogaster. Science 287, 2185–2195.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.
Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, et al., 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330, 86–88.
Badisco, L., Huybrechts, J., Simonet, G., Verlinden, H., Marchal, E., et al., 2011. Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. PLoS ONE 6 (3), e17274.
Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.
Brezden, B.L., Yeoman, M.S., Gardner, D.R., Benjamin, P.R., 1999. FMRFamide-activated Ca2+ channels in Lymnaea heart cells are modulated by "SEEPLY," a neuropeptide encoded on the same gene. J. Neurophysiol. 81, 1818–1826.
Brody, T., Cravchik, A., 2000. Drosophila melanogaster G protein-coupled receptors. J. Cell. Biol. 150, F83–F88.
Dotson, E.M., Beard, C.B., 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, Triatoma dimidiata. Insect Mol. Biol. 10, 205–215.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Fan, Y., Sun, P., Wang, Y., He, X., Deng, X., Chen, X., Zhang, G., Chen, X., Zhou, N., 2010. The G protein-coupled receptors in the silkworm, Bombyx mori. Insect Biochem. Mol. Biol. 40, 581–591.
Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36 (10), 3420–3435.
Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., Grimmelikhuijzen, C.J.P., 2006. A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. Prog. Neurobiol. 80, 1–19.
Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., Grimmelikhuijzen, C.J.P., 2008. A genome-wide inventory of neurohormone GPCRs in the red flour beetle Tribolium castaneum. Front. Neuroendocrinol. 29, 142–165.
Hewes, R.S., Taghert, P.H., 2001. Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome Res. 11, 1126–1142.
Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M., Zwiebel, L.J., 2002. G protein-coupled receptors in Anopheles gambiae. Science 298, 176–178.
Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., 2002. The genome sequence of the Malaria Mosquito *Anopheles gambiae*. Science 298, 129–149.
International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8 (2), e1000313.
Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., et al., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. U.S.A. 107, 12168–12173.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Li, C., Yun, X., Hu, X., Zhang, Y., Sang, M., Liu, X., Wu, W., Li, B., 2013. Identification of G protein-coupled receptors in the pea aphid, Acyrthosiphon pisum. Genomics 102, 345–354.
Nässel, D.R., 2002. Neuropeptides in the nervous system of Drosophila and other insects: multiple roles as neuromodulators and neurohormones. Prog. Neurobiol. 68, 1–84.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., et al., 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316, 1718–1723.
Ons, S., Richter, F., Urlaub, H., Pomar, R.R., 2009. The neuropeptidome of Rhodnius prolixus brain. Proteomics 9, 788–792.
Ons, S., Sterkel, M., Diambra, L., Urlaub, H., Rivera-Pomar, R., 2011. Neuropeptide precursor gene discovery in the Chagas disease vector Rhodnius prolixus. Insect Mol. Biol. 20, 29–44.
Orchard, I., Lee, D.H., da Silva, R., Lange, A.B., 2011. The proctolin gene and biological effects of proctolin in the blood-feeding bug, *Rhodnius prolixus*. Front. Endocrinol. (Lausanne) 2, 59.
Paluzzi, J.P., O'Donnell, M.J., 2012. Identification, spatial expression analysis and functional characterization of a pyrokinin-1 receptor in the Chagas' disease vector, Rhodnius prolixus. Mol. Cell. Endocrinol. 363, 36–45.
Paluzzi, J.P., Park, Y., Nachman, R.J., Orchard, I., 2010. Isolation, expression analysis, and functional characterization of the first antidiuretic hormone receptor in insects. Proc. Natl. Acad. Sci. U.S.A. 107, 10290–10295.
Paluzzi, J.P., Haddad, A.S., Sedra, L., Orchard, I., Lange, A., 2014. Functional characterization and expression analysis of the myoinhibiting peptide receptor in the Chagas disease vector, *Rhodnius prolixus*. Mol. Cell. Endocrinol. 399, 143.
Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061–1067.
Rassi, A.J., Rassi, A., Marín-Neto, J.A., 2010. Chagas disease. Lancet 375, 1388–1402.
Ribeiro, J.M., Genta, F.A., Sorgine, M.H., Logullo, R., Mesquita, R.D., et al., 2014. An insight into the transcriptome of the digestive tract of the bloodsucking bug, Rhodnius prolixus. PLoS NTD 8, e2594.
Salcedo, E., Huber, A., Henrich, S., Chadwell, L.V., Chou, W.H., Paulsen, R., Britt, S.G., 1999. Blue- and green-absorbing visual pigments of Drosophila: ectopic expression and physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6 rhodopsins. J. Neurosci. 19, 10716–10726.
Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092.
Sterkel, M., Urlaub, H., Rivera-Pomar, R., Ons, S., 2011. Functional proteomics of neuropeptidome dynamics during the feeding process of *Rhodnius prolixus*. J. Prot. Res. 10, 3363–3371.
Sterkel, M., Oliveira, P.L., Urlaub, H., Hernandez-Martinez, S., Rivera-Pomar, R., Ons, S., 2012. OKB, a novel family of brain-gut neuropeptides from insects. Ins. Biochem. Mol. Biol. 42, 466–473.
Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., Shinoda, T., 2014. Transcriptome analysis of neuropeptides and G-protein coupled receptors (GPCRs) for neuropeptides in the brown planthopper Nilaparvata lugens. Peptides 53, 125–133.
Vafopoulou, X., Steel, C.G.H., 2014. Synergistic induction of the clock protein PERIOD by insulin-like peptide and prothoracicotropic hormone in Rhodnius prolixus (Hemiptera): implications for convergence of hormone signaling pathways. Front. Physiol. 5, 41.
Vafopoulou, X., Steel, C.G.H., Terry, K.L., 2007. Neuroanatomical relations of prothoracicotropic hormone neurons with the circadian timekeeping system in the brain of larval and adult Rhodnius prolixus (Hemiptera). J. Comp. Neurol. 503, 511–524.
Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Arch. Insect Biochem. Physiol. 43, 49–63.
Velarde, R.A., Sauer, C.D., Walden, K.K.O., Fahrbach, S.E., Robertson, H.M., 2005. Pteropsin: a vertebrate-like non-visual opsin expressed in the honey bee brain. Insect Biochem. Mol. Biol. 35, 1367–1377.
Wegener, C., Gorbashov, A., 2008. Molecular evolution of neuropeptides in the genus Drosophila. Genome Biol. 9, R131.
Wheeler, M.M., Ament, S.A., Rodriguez-Zas, S.L., Robinson, G.E., 2013. Brain gene expression changes elicited by peripheral vitellogenin knockdown in the honey bee. Insect Mol. Biol. 22, 562–573.
Zandawala, M., Paluzzi, J., Orchard, I., 2010. Isolation and characterization of the cDNA encoding DH31 in the kissing bug; Rhodnius prolixus. Mol. Cell. Endocrinol. 1–10.
Zandawala, M., Li, S., Hauser, F., Grimmelikhuijzen, C.J.P., Orchard, I., 2013. Isolation and functional characterization of calcitonin-like diuretic hormone receptors in Rhodnius prolixus. PLoS One 8, e82466.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhang, Z., Peng, Z.-Y., Yi, K., Cheng, Y., Xia, Y., 2012. Identification of representative genes of the central nervous system of the locust, Locusta migratoria manilensis by deep sequencing. J. Insect Sci. 12, 86.

Please cite this article in press as: Ons, S., et al., Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis, Insect Biochemistry and Molecular Biology (2015), http://dx.doi.org/10.1016/j.ibmb.2015.05.003