

2

**The hormone neuroparsin seems essential in Lepidoptera but not in  
4 domesticated silkworms**

6

8 Jan A. Veenstra<sup>1</sup>,

10 <sup>1</sup> INCIA UMR 5287 CNRS, University of Bordeaux, Pessac, France.

12 Corresponding Author:

Jan A. Veenstra<sup>1</sup>

14 INCIA UMR 5287 CNRS, Université de Bordeaux, allée Geoffroy St Hillaire, CS 50023, 33 615  
Pessac Cedex, France

16 Email address: [jan-adrianus.veenstra@u-bordeaux.fr](mailto:jan-adrianus.veenstra@u-bordeaux.fr)

## Abstract

18

The primary sequence of the Arthropod neurohormone neuroparsin is so variable that so far no  
20 orthologs from moths and butterflies have been characterized, even though classical neurosecretory  
stains identify cells that are homologous to those producing this hormone in other insect species. Here  
22 Lepidopteran cDNAs showing limited sequence similarity to other insect neuroparsins are described.  
That these cDNAs do indeed code for authentic neuroparsins was confirmed by *in situ* hybridization in  
24 the wax moth, *Galleria mellonella*, which labeled the neuroparsin neuroendocrine cells. Although in  
virtually all genome assemblies from Lepidoptera a neuroparsin gene could be identified, the genome  
26 assembly from the silkworm, *Bombyx mori*, has a neuroparsin gene containing a 16 nucleotide deletion  
that renders this gene nonfunctional. Although only a small number of all silkworm strains carry this  
28 deletion, it suggests that the domestication of the silkworm has rendered the function of this  
neurohormone dispensable.

30

32

34 **Key words:** Neuroparsin – venus kinase receptor – Lepidoptera – *Bombyx* – silkworm – domestication

## 1. Introduction

36

In a 1928 publication Ernst Scharer described what are now known as neuroendocrine cells  
38 producing vasopressin and oxytocin (Scharer, 1928). At the time the concept of what was later called  
neurosecretion and neuroendocrinology, but what at the time he called internal secretion (“innere  
40 Sekretion”), was not readily accepted by vertebrate physiologists. However, insect physiologists had  
already demonstrated the existence of a “brain hormone” that activates the prothoracic gland (*e.g.*  
42 Kopeć, 1922) and this strongly stimulated research on neurosecretory cells in insects. One of the most  
popular staining methods uses paraldehyde fuchsin, which stains several cells in the insect *pars*  
44 *intercerebralis* that have axons projecting to the *corpora cardiaca*, where they release their products  
into the hemolymph. There are generally two different paraldehyde fuchsin cell types in the *pars*  
46 *intercerebralis*, which differ by their staining intensity. We now know that the less intensely staining  
cells produce insect insulins and the other neuroparsins (Goltzené et al., 1992).

48 The structure of neuroparsin was first identified from *Locusta migratoria* (Girardie et al, 1989).  
Nevertheless, relatively little is known about the physiological relevance of this hormone in insects, no  
50 doubt due in large part to its absence from *Drosophila melanogaster* (Veenstra, 2010), the insect  
neuropeptide model par excellence. In *Locusta migratoria* the hormone antagonizes juvenile hormone.  
52 It delays vitellogenesis and immunoneutralization of neuroparsin accelerates vitellogenesis and induces  
green pigmentation. These effects are similar to those of juvenile hormone (Girardie et al., 1987). Once  
54 its cDNA was cloned, neuroparsin was found to be expressed also outside the central nervous system  
(Lagueux et al., 1992) and the identification of several distinct cDNAs showed that they were derived  
56 from a single gene that is alternatively spliced. The genome of *L. migratoria* revealed the neuroparsin  
gene to have seven coding exons that allow the production of five different neuroparsin transcripts  
58 (Veenstra, 2014). Work on another migratory locust, *Schistocerca gregaria*, revealed a very similar  
situation, although the genome for this species has not yet been sequenced and so far only four different  
60 neuroparsin transcripts have been identified (Girardie et al., 1998; Janssen et al., 2001; Claeys et al.,  
2003, 2005, 2006a,b). Interestingly, expression of the different transcripts is significantly different  
62 between the solitary and gregarious phases in the latter species (Claeys et al., 2006a).

The extensive alternative splicing of neuroparsin transcripts seems to be limited to migratory  
64 locusts, as to my knowledge it has not been found in other Arthropods. Some species, notably various  
Decapods (Veenstra, 2016) and blood sucking bugs like *Rhodnius prolixus* (KQ034340.1) and *Cimex*

66 *lectularis* (Predel et al., 2018), however have multiple neuroparsin genes. In this respect neuroparsin  
resembles insulin, which in many species also has multiple genes. This is not the only similarity  
68 between neuroparsin and insulin. Thus, both neuroparsin and insulin promote neurite outgrowth in  
locust brain (Vanhems et al., 1990). Furthermore, identification of ovarian ecdysteroid hormone from  
70 the mosquito *Aedes aegypti* revealed it to be a neuroparsin ortholog (Brown et al., 1998) and the  
subsequent identification of a venus kinase as its receptor (Vogel et al., 2015) showed that insulin and  
72 neuroparsin use similar receptors in insects. In mosquitoes both hormones stimulate the ovary to  
produce ecdysteroids which in turn trigger vitellogenin synthesis by the fat body (Dhara et al., 2003).  
74 Thus these two hormones seem somewhat complimentary. This is also suggested by venus kinase  
RNAi injections in *S. gregaria*, where both neuroparsin and insulin transcripts are upregulated .  
76 Downregulation of the venus kinase receptor by RNAi leads to a delay in vitellogenin synthesis and  
strongly inhibits production of ecdysone and several of the enzymes needed for its synthesis.  
78 Nevertheless, overall reproductive success appeared to be only slightly impacted in *Schistocerca*  
(Lenaerts et al., 2017).

80 The brains of Lepidoptera also typically contains two cell types that are stained by paraldehyde  
fuchsin (Panov and Kind, 1963) and Lepidopteran insulins have been extensively studied in the  
82 silkworm (for review see: Mizoguchi and Okamoto, 2013). However, in spite of much research on  
neuropeptides in the silkworm, and the presence of two venus tyrosine kinase receptors in its genome  
84 (Vanderstraete et al., 2013), their putative ligands have so far escaped identification. I here show that  
neuroparsin seems ubiquitously present in Lepidoptera but that in some domesticated silkworm strains  
86 the gene has mutated in such fashion that it can no longer be released as a hormone into the  
hemolymph.

88

## 2. Materials and Methods

90

### 2.1. Bioinformatics

92 After the initial identification of the *Galleria* and *Bombyx* neuroparsin genes 63 lepidopteran genome  
assemblies available at NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=lepidoptera>) were  
94 analyzed for the presence of orthologous genes. Predicted neuroparsin precursors were deduced and the  
size of the intron between the first and second coding exons was determined. In most cases the three  
96 coding exons were found on the same contig, but in several genome assemblies the contigs are very

small and in those cases the the coding exons were connected based on homology to produce a  
98 neuroparsin precursor. In four assemblies no neuroparsin gene was found, in two of these species,  
*Heliconius hecale* and *Heliothis virescens*, the genes were assembled from individual reads recovered  
100 from genomic SRAs, SRR7162651 and SRR5463746 respectively, using the sratoolkit  
([www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/](http://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/)) and Trinity (Grabherr et al., 2011) using a method  
102 described previously (Veenstra, 2019).

## 104 2.2 *In situ hybridization*

The *in situ* hybridization protocol is based on a protocol provided by Dušan Žitňan for *Drosophila* as  
106 described *e.g.* here (Kim et al., 2006). Dissections were done in 0.9 % NaCl and tissues were fixed in  
phosphate buffered 4% paraformaldehyde for 2 hrs at room temperture in Eppendorf tubes. Tissues  
108 were subsequently washed thrice for 30 min in PBS with 0.2 % Tween 20 (PBST) and once for 30 min  
in 70% ethanol. After replacing the 70% ethanol the tissues were stored at -20 °C for three to five days.  
110 Tissues were next washed thrice for 25 min in PBST and then incubated with Proteinase K (50 µg/ml)  
in PBST for 30 min. The proteinase K was stopped by washing the tissue for 30 min in glycine (2  
112 mg/ml) in PBST, followed by two washes of 20 min in PBST. Next tissues were fixed a second time in  
phosphate buffered 4% paraformaldehyde for 1 hr followed by two washes in PBST for 20 min each.  
114 This was followed by incubations for 20 min in 1:1 mixture of hybridization solution (HS: 50 µg/ml  
heparin, 100 µg/ml salmon testes DNA, 750 mM NaCl, 75 mM sodium citrate, in deionized RNase free  
116 water, pH 7.0, 50% deionized formamide, 0.1 % Tween 20) and PBST, followed by 25 min in HS, both  
at roomtemperature. HS was then replaced with fresh HS and the tissues were incubated for 3 hrs at 48  
118 °C, after which HS was replaced with fresh HS containing 10% of digoxigenin hybridization probe  
which had previously been brought to 95 °C for either 3 min, in case of a previously used probe, or 45  
120 min when a hybridization probe is used for the first time. Hybridization was carried out overnight at 48  
°C. The following day hybridization probe was recovered and stored at -20 °C for future use. Tissues  
122 are washed in fresh HS four times, thrice for 2 to 2.5 hrs, and then overnight, all washes in HS were  
done at 48 °C. The next day tissues were brought to room temperature and washed in a 1:1 mixture of  
124 PBST and HS, follwed by three 20 min washes in PBST at room temperature. Next tissues were  
saturated with 1% BSA in PBST for 1 hr. The 1% BSA in PBST was then used to dilute sheep anti-  
126 digoxigenin Fab-fragments conjugated to alkaline phosphatase (1:1000, Sigma-Aldrich) in which the  
tissues were incubated overnight at room temperature. Next morning tissues were washed thrice 20 min

128 with PBST, followed by three washes in freshly prepared alkaline phosphate buffer (APB, 100 mM  
Tris, 50 mM MgCl<sub>2</sub>, 100 mM NaCl, pH 9.5 and 0.1% Tween 20). During the last wash, the tissues  
130 were transferred to small glass dissection dishes and then the location of the digoxigenin probe was  
visualized by replacing the wash solution with APB containing 20 µl/ml of a commercial NBT/BCIP  
132 stock solution (Sigma-Aldrich, 18.75 mg/ml nitro blue tetrazolium chloride and 9.4 mg/ml 5-bromo-4-  
chloro-3-indolyl-phosphate, toluidine-salt in 67% DMSO). Color development was followed under a  
134 dissecting microscope and took about five minutes. Once color development was judged satisfactory,  
the alkaline phosphatase was stopped by changing the staining solution with 100 mM phosphate  
136 buffer, pH 7.0. This was followed by four 15 min washes in increasing concentrations of glycerol (20  
%, 40 %, 60 % and 80 %) and finally tissues were mounted between a slide and a coverslip in 80 %  
138 glycerol. All incubations at room temperature were performed on gently rotating orbital shaker.

#### 140 2.3. Hybridization probes

RNA was extracted and reverse transcribed using random primers and M-MuLV reverse transcriptase  
142 (New England Biolabs) and a PCR product of about 300 nucleotides from a coding sequence was then  
amplified by PCR using Q5® High-Fidelity DNA polymerase (New England Biolabs). The resulting  
144 PCR products were gel purified, quantified and their identities as authentic neuroparsin cDNAs  
confirmed by automated Sanger sequencing. Sense and anti-sense digoxigenin labeled probes were  
146 then produced using Taq polymerase (New England Biolabs) and 40 cycles of PCR with either the  
sense or anti-sense primer in a Taq polymerase PCR mixture to which 0.067 mM of Digoxigenin-5-  
148 aminoallyl-dUTP (Jena Bioscience GmbH, Jena, Germany) had been added and in which the  
concentration of dTTP had been lowered to 0.133 mM. The final PCR product was used directly in the  
150 *in situ* hybridization procedure.

#### 152 2.4. Primer sequences used

##### *Galeria* neuroparsin

154 forward primer 5'-CGAGTGTCCAGTATGCATCG-3'

reverse primer 5'-TTGGGATTAAGGAGCCATTG-3'

##### 156 *Bombyx* neuroparsin

forward primer 5'-GAGAGACGAACCGGAAATCA-3'

158 reverse primer 5'-CCATGCCTTGACAGATACCC-3'

160 2.5. *Insects*

*Galleria mellonella* larvae were purchased from St. Laurent (La Chapelle Saint Laurent, France), while *Bombyx mori* pupae were a generous gift from Michal Žurovec from the University of Southern Bohemia.

164

**3. Results**

166 Extensive use of the BLAST program led to the identification of a neuroparsin precursor candidate that had very limited sequence homology with its orthologs from other species (Fig. 1), but that within the Lepidoptera is relatively well conserved (Fig. 2 and Table S1). It typically has a signal peptide and analysis of a limited number of RNAseq SRAs from *B. mori* (Table S2) suggested it to be relatively specific for the brain.

Several of the Lepidopteran genome assemblies are little more than first drafts and it is therefore not surprising that I was unable to find evidence for a neuroparsin gene in *Operophtera brumata*. On the other hand there is abundant genomic and transcriptome data for *Plutella xylostella* (see e.g. <http://dbm.dna.affrc.go.jp/px/>). This species is one of two studied here that does not belong to the Apoditrysia taxon. The other one is *Eumeta japonica*, and its neuroparsin precursor sequence is very different from that of the Apoditrysia (Fig. 1). Hence it seems plausible that the *Plutella* neuroparsin sequence was not recognized because it is too different from that of the other Lepidoptera.

178 To confirm or invalidate the hypothesis that these sequences might represent neuroparsin orthologs, whole mount *in situ* hybridization of brains was performed in *Galleria mellonella* larvae as well as adults. RNA extraction of *Galleria* heads and its subsequent reverse transcription allowed the production of a neuroparsin PCR product the identity of which was confirmed by Sanger sequencing (Figs. S1,S2). In both larvae and adults *in situ* hybridization revealed two bilateral groups of four large cells in the *pars intercerebralis* (Fig. 3) in the location where neuroparsin expressing cells were expected based on paraldehyde fuchsin staining reports (Panov and Kind, 1963). RNA extraction of a single *Bombyx* brain and its subsequent reverse transcription similarly allowed the production of a neuroparsin PCR product that was sequenced and found to have the 16 base pair solution (Figs. S3,S4). Although *in situ* hybridization procedures were identical to those performed on *Galleria*, which were run in parallel, no *in situ* hybridization signal was obtained in *Bombyx* adult brains. This could be due to low expression levels of neuroparsin in *B. mori*, as suggested by the 0.00 TPM (transcripts per

190 million) values for the two brain RNAseq transcripts present at silkbase ([http://silkbase.ab.a.u-](http://silkbase.ab.a.u-tokyo.ac.jp)  
192 [tokyo.ac.jp](http://silkbase.ab.a.u-tokyo.ac.jp)). Digoxigenin probes generated with the sense probes did not yield any signal in either  
species.

Intriguingly in the *Bombyx mori* genome assembly, the neuroparsin gene does not code for a  
194 signal peptide. However, a signal peptide is predicted from the orthologous *B. mandarina* gene (Fig. 4),  
the species from which *B. mori* was domesticated about 5,000 years ago (Sun et al., 2012; Chen et al.,  
196 2019). There are two significant differences between the two genes. The first one concerns the intron  
between the first and second coding exons. In the domesticated silkworm genome assembly it contains  
198 9,493 but that of the wild silkworm has only 827 base pairs. The second and more important difference  
is a unique deletion of 16 base pairs from the first coding exon (Fig. 4). Analysis of *B. mori* brain  
200 RNAseq SRAs yielded a total of 4,358 spots and Trinity unambiguously confirmed the proposed  
mRNA. This mRNA contains the unique deletion, as was expected since these RNAseq SRAs are all  
202 obtained from the same strain as used for genome sequencing.

The deletion of 16 base pairs from the first coding exon either leads to a very small peptide with  
204 a signal sequence or, in case the first three methionine codons were to be ignored, to a protein that  
contains the neuroparsin sequence, but lacks a signal peptide (Fig. 4). Therefore this mutation makes it  
206 impossible for the neuroendocrine cells expressing this mutated neuroparsin gene to secrete this  
neurohormone. Over time a large number of different silkworm strains have been sequenced, in order  
208 to analyze the domestication of the silkworm and its underlying evolutionary process (Xia et al.,  
2009; Xiang et al., 2018). Sequences for still other lines are available at silkbase ([http://silkbase.ab.a.u-](http://silkbase.ab.a.u-tokyo.ac.jp)  
210 [tokyo.ac.jp](http://silkbase.ab.a.u-tokyo.ac.jp)). Although the latest *B. mori* genome assembly (Kawamoto et al., 2019) shows a 16  
basepair deletion in the neuroparsin coding sequence, only a limited number of silkworm strains have  
212 this deletion. Thus in the 161 different *B. mori* strains in which this could be determined, there were  
only twelve that carried the 16 base pair deletion while four other strains had both alleles. As genome  
214 assemblies are all based on single individuals, the actual number of *B. mori* strains where this allele is  
present is almost certainly larger, but it must be a minority. These numbers are not consistent with the  
216 Hardy-Weinberg equilibrium and, hence, in most populations the allele will be fixed.

It seemed interesting to know whether the enormous increase in size of the intron between the  
218 first and second coding exons might be of functional significance. I therefore determined, where  
possible, the size of this intron in other Lepidopteran genomes, as well as the various *Bombyx* strains.  
220 Whereas it was relatively easy to determine coding sequences in genome assemblies with large



222 numbers of small contigs, the determination of the size of this intron is more difficult. Nevertheless,  
there were only a few species in which the size of this intron could not be determined. In the large  
majority of species, this intron is quite small, and there are only a few species where it is larger than  
224 1500 base pairs; the largest intron was found in *Tuta absoluta* (2285, Table S1). Determining the size of  
the intron in all the *Bombyx* strains was more difficult, in five *B. mandarina* strains where this could be  
226 determined it varied between 827 and 1292. In the thirty *B. mori* strains where this could be established  
it varied between 918 and 1298; none of these strains had the 16 base pair deletion. There was not a  
228 single *B. mori* strain having this deletion for which it was possible to determine the size of the intron,  
apart from the of the official sequence assembly of the p50T strain (Kawamoto et al., 2019).  
230 Nevertheless, in several strains its size had to be at least more than 1900 base pairs and this was found  
both in strains that have the 16 basepair deletion and strains that do not. Although this suggests that the  
232 intron size increased before the deletion occurred, it does not definitively prove it.

#### 234 4. Discussion

236 *Bombyx mori* has been and still is the model for lepidopteran neuropeptides and it was  
surprising that neuroparsin that seemed to be produced in large neuroendocrine cells of the brain had  
238 escaped detection in this species as well as other Lepidoptera. It is now clear that Lepidoptera do have  
such a hormone, but its sequence is so different from that of other Arthropod neuroparsins, that it is  
240 difficult to recognize as such. Nevertheless, the expression in *Galleria* brain neuroendocrine cells of  
this protein confirms that it is indeed neuroparsin.

242 In the *Bombyx* strain used preferentially in the laboratory and the one with the best genome  
assembly (Kawamoto et al., 2019) the neuroparsin gene shows two mutations. The first one makes it  
244 impossible for the putative neuroparsin precursor to have a signal peptide and hence it can not be  
released as a hormone into the hemolymph. Although, it is also possible that no neuroparsin is made at  
246 all, and that only small peptide with a signal peptide is the protein product made from this gene. Either  
way, this probably explains why neuroparsin has not previously been identified from this species.

248 It is interesting to note, that there is second mutation in the *Bombyx* neuroparsin gene, which  
concerns a very significant increase in the size of the intron between the first and second coding exons.  
250 In those moths and butterflies where its size could be determined it was never more than 2285, and in  
most of them it was less than 1000 base pairs. Yet in the sequenced *B. mori* strain it is 9493 base pairs

252 long. It is noticeable that neuropeptide genes that are heavily expressed in small number of neurons,  
such as the insulins, AKHs, or the *Locusta* vasopressin genes, all seem to have very small introns,  
254 suggesting that small intron sizes may facilitate efficient neuropeptide production. It seemed possible  
that once a 16 base pair deletion effectively inactivated the neuroparsin gene there was no longer any  
256 selection for maintaining the small size of this intron. Alternatively, if for some unknown reason the  
neuroparsin gene has no function in domesticated silkworms, the increase of the size of this intron may  
258 have preceded the deletion. Results suggested that the deletion in the coding sequence occurred after  
the increase of the intron size. If one accepts the notion that small intron sizes facilitate efficient  
260 expression of (neuropeptide) genes, it then also suggests that during domestication the importance of  
the neuroparsin gene was diminished well before the gene was completely inactivated by the deletion  
262 in its coding region. Such a scenario may explain the absence of a detectable *in situ* hybridization  
signal in *B. mori*.

264 It is indeed surprising, that while in virtually all Lepidopteran genomes a neuroparsin gene can  
be detected that seems to be functional, it is only in some domesticated *Bombyx* strains that this gene  
266 has apparently become superfluous. It is well documented that domestication leads to significant  
changes in the genome of a species and this also occurred in the silkworm (Xiang et al., 2018). As this  
268 mutated gene is only present in some but not all domesticated silkworms it is clearly not a mutation that  
facilitated domestication, it is rather a gene that domesticated silkworms can do without, while  
270 remaining essential in wild moths.

It is obvious that just like *Drosophila melanogaster* the domesticated silkworm survives well  
272 without a functional neuroparsin gene. It can thus not be excluded that it is a pure coincidence that this  
happened in some domesticated strains of *Bombyx mori* and not in other Lepidoptera, assuming that the  
274 absence of the gene in *Plutella xylostella* and *Operophtera brumata* is indeed due to technical problems  
rather than a genuine absence from their genomes.

276 A recent publication on *Anopheles coluzzii* showed the venus kinase receptor to be required for  
protection against infection by *Plasmodium* parasites (Gouignard et al., 2019). As the venus kinase  
278 receptor can also be activated by L-amino acids (*e.g.* Ahier et al., 2009), this does not necessarily show  
the implication of neuroparsin. Nevertheless, within in this context it is at least intriguing to note that  
280 some of the species with the largest intron sizes all have some type of natural exterior protection during  
the larval stages which not only makes chemical control of such species difficult, but likely also  
282 protects them against viruses, bacteria and perhaps parasites. Thus *Cydia pomonella* (intron size 1598,

Table S1) larvae live and feed inside apples, larvae of *Galleria mellonella* (intron size 1884, Table S1)  
284 live inside bee hives, *Tuta absoluta* (intron size 2285, Table S1) is a leaf miner and caterpillars of  
*Megathymus ursus* (intron size 1549, Table S1) feed inside yucca roots and agave leaves (Cong et al.,  
286 2019). In any case, a hormone that would be important for protection against parasites or infections,  
would likely be less important in a domesticated species, as diseased animals would be rapidly  
288 eliminated by silk farmers.

The identification of Lepidopteran neuroparsins shows how variable its structure can be in  
290 insects and it illustrates the difficulty of finding putative orthologs of this hormone in other Arthropods.  
Although not limited to insects as the Decapod neuroparsins convincingly show, it remains unknown  
292 whether this hormone is present in other Arthropods like *e.g.* Chelicerates. Venus kinase receptors have  
a wide distribution within Protostomes and even Deuterostomes (Vanderstraete et al., 2013; Dissous et  
294 al., 2014). Since neuroparsin can activate this type of receptor (Vogel et al., 2015), *a priori*, one might  
expect neuroparsin orthologs to have a similar wide distribution. However, these receptors are also  
296 activated by amino acids (*e.g.* Ahier et al., 2009; Vanderstraete et al., 2014; Gougnard et al., 2019) and  
hence the distribution of neuroparsins could be much more limited. It is therefore regrettable that the  
298 extreme variability of this hormone in insects suggests, that it will be very difficult to identify  
neuroparsin orthologs, if they even exist.

300  
302 **Acknowledgements.** I thank Michal Žurovec for *Bombyx mori* pupae and Dušan Žitňan for his *in situ*  
hybridization protocol.

## 304 **References**

306 Ahier, A., Rondard, P., Gougnard, N., Khayath, N., Huang, S., Trolet, J., Donoghue, D.J., Gauthier, M.,  
Pin, J.P., Dissous, C. 2009. A new family of receptor tyrosine kinases with a venus flytrap binding  
308 domain in insects and other invertebrates activated by aminoacids. PLoS One 4:e5651. doi:  
10.1371/journal.pone.0005651.

310 Brown, M.R., Graf, R., Swiderek, K.M., Fendley, D., Stracker, T.H., Champagne, D.E., Lea, A.O.  
312 1998. Identification of a steroidogenic neurohormone in female mosquitoes. J. Biol. Chem. 273:3967-  
3971. doi:10.1371/journal.pone.0017274.

314 Chen, D.B., Zhang, R.S., Bian, H., Li, Q., Xia, RX., Li, YP., Liu YQ., Lu, C.. 2019. Comparative  
316 mitochondrial genomes provide new insights into the true wild progenitor and origin of domestic  
silkworm *Bombyx mori*. Int. J. Biol. Macromol. 131:176-183. doi: 10.1016/j.ijbiomac.2019.03.002.

318

- 320 Claeys, I., Simonet, G., Van Loy, T., De Loof, A., Vanden Broeck, J. 2003. cDNA cloning and transcript  
distribution of two novel members of the neuroparsin family in the desert locust, *Schistocerca gregaria*.  
Insect Mol. Biol. 12:473-481. doi:10.1046/j.1365-2583.2003.00431.x.
- 322 Claeys, I., Simonet, G., Breugelmans, B., Van Soest, S., Franssens, V., Sas, F., De Loof, A., Vanden  
324 Broeck, J. 2005. Quantitative real-time RT-PCR analysis in desert locusts reveals phase dependent  
differences in neuroparsin transcript levels. Insect Mol. Biol. 14:415-422. doi:10.1111/j.1365-  
326 2583.2005.00572.x.
- 328 Claeys, I., Breugelmans, B., Simonet, G., Franssens, V., Van Soest, S., Vanden Broeck J. 2006a.  
Regulation of *Schistocerca gregaria* neuroparsin transcript levels by juvenile hormone and 20-  
330 hydroxyecdysone. Arch. Insect Biochem. Physiol. 62:107-115. doi:10.1002/arch.20127.
- 332 Claeys, I., Breugelmans, B., Simonet, G., Van Soest, S., Sas, F., De Loof, A., Vanden Broeck, J. 2006b.  
Neuroparsin transcripts as molecular markers in the process of desert locust (*Schistocerca gregaria*)  
334 phase transition. Biochem. Biophys. Res. Commun. 341:599-606. doi:10.1016/j.bbrc.2006.01.011.
- 336 Cong, Q., Li, W., Borek, D., Otwinowski, Z., Grishin, N.V. 2019. The Bear Giant-Skipper genome  
suggests genetic adaptations to living inside yucca roots. Mol. Genet. Genomics. 2019 Feb;294(1):211-  
338 226. doi: 10.1007/s00438-018-1494-6.
- 340 Dhara, A, Eum, J.H, Robertson, A., Gulia-Nuss, M., Vogel, K.J., Clark, K.D., Graf, R., Brown, M.R.,  
Strand, M.R. 2013. Ovary ecdysteroidogenic hormone functions independently of the insulin receptor  
342 in the yellow fever mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 43:1100-1108. doi:  
10.1016/j.ibmb.2013.09.004.
- 344 Dissous, C, Morel, M, Vanderstraete, M., 2014. Venus kinase receptors: prospects in signaling and  
346 biological functions of these invertebrate kinases. Front. Endocrinol. 5:72. doi:  
10.3389/fendo.2014.00072.
- 348 Girardie, J., Bourême, D., Couillaud, F., Tamarelle, M., Girardie, A., 1987. Anti-juvenile hormone  
350 effect of neuroparsin A, a neuroprotein isolated from locust corpora cardiaca. Insect Biochem. 17, 977-  
983. doi:10.1016/0020-1790(87)90106-5.
- 352 Girardie, J., Girardie, A., Huet, J.C., Pernollet, J.C. 1989. Amino acid sequence of locust neuroparsins.  
354 FEBS Lett. 245:4-8. doi:10.1016/0014-5793(89)80179-6.
- 356 Girardie, J., Huet, J.C., Atay-Kadiri, Z., Ettaouil, S., Delbecque, J.P., Fournier, B., Pernollet, J.C.,  
Girardie, A. 1998. Isolation, sequence determination, physical and physiological characterization of the  
358 neuroparsins and ovary maturing parsins of *Schistocerca gregaria*. Insect Biochem. Mol. Biol. 28:641-  
650. doi:10.1016/S0965-1748(98)00053-8.
- 360 Goltzené, F., Holder, F., Charlet, M., Meister, M., Oka, T. 1992. Immunocytochemical localization of  
362 Bombyx-PTTH-like molecules in neurosecretory cells of the brain of the migratory locust, *Locusta*  
*migratoria*. A comparison with neuroparsin and insulin-related peptide. Cell Tissue Res. 269:133-140.  
364 doi:10.1007/BF00384733.

- 366 Gouignard, N., Cherrier, F., Brito-Fravallo, E., Pain, A., Zmarlak, N.M., Cailliau, K., Genève, C.,  
368 Vernick, K.D., Dissous, C., Mitri, C. 2019. Dual role of the *Anopheles coluzzii* Venus Kinase Receptor  
in both larval growth and immunity. *Sci. Rep.* 9:3615. doi: 10.1038/s41598-019-40407-x.
- 370 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L.,  
372 Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F.,  
Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. 2011. Full-length transcriptome  
374 assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644-652.
- Janssen, T., Claeys, I., Simonet, G., De Loof, A., Girardie, J., Vanden Broeck, J. 2001. cDNA cloning  
376 and transcript distribution of two different neuroparsin precursors in the desert locust, *Schistocerca*  
*gregaria*. *Insect Mol. Biol.* 10:183-189. doi:10.1046/j.1365-2583.2003.00431.x.
- 378 Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., Fujiyama, A., Kiuchi,  
380 T., Yamamoto, K., Shimada, T. 2019. High-quality genome assembly of the silkworm, *Bombyx mori*.  
*Insect Biochem. Mol. Biol.* 107:53-62. doi: 10.1016/j.ibmb.2019.02.002.
- 382 Kim, Y.J., Žitňan, D., Galizia, C.G., Cho, K.H., Adams, M.E., 2006. A command chemical triggers an  
384 innate behavior by sequential activation of multiple peptidergic ensembles. *Curr. Biol.* 16:1395–  
1407. doi: 10.1016/j.cub.2006.06.027.
- 386 Kopeć, S., 1922. Studies on the necessity of the brain for the inception of insect metamorphosis. *Biol.*  
388 *Bull.* 42: 323–342. doi:10.2307/1536759.
- 390 Lagueux, M., Kromer, E., Girardie, J. 1992. Cloning of a *Locusta* cDNA encoding neuroparsin A.  
*Insect Biochem. Mol. Biol.* 22:511-516. doi:10.1016/0965-1748(92)90027-C.
- 392 Lenaerts, C., Palmans, J., Marchal, E., Verdonck, R., Vanden Broeck, J. 2017. Role of the venus kinase  
394 receptor in the female reproductive physiology of the desert locust, *Schistocerca gregaria*. *Sci. Rep.*  
7:11730. doi: 10.1038/s41598-017-11434-3.
- 396 Mizoguchi, A., Okamoto, N. 2013. Insulin-like and IGF-like peptides in the silkworm *Bombyx mori*:  
398 discovery, structure, secretion, and function. *Front. Physiol.* 4: 217. doi: 10.3389/fphys.2013.00217.
- 400 Панов, А.А., Кинд, Т.В. , 1963. Система нейросекреторных клеток головного мозга  
402 чешуекрылых (Lepidoptera, Insecta). Докл. Акад. наук СССР 153:1186-1189.
- Predel, R., Neupert, S., Derst, C., Reinhardt, K., Wegener, C. 2018. Neuropeptidomics of the bed bug  
404 *Cimex lectularius*. *J. Proteome Res.* 17:440-454. doi: 10.1021/acs.jproteome.7b00630.
- 406 Scharrer, E. 1928. Die Lichtempfindlichkeit blinder Elritzen. *Z. vergl. Physiol.* 7:1-38.
- 408 Sun, W., Yu, H., Shen, Y., Banno, Y., Xiang, Z., Zhang, Z. 2012. Phylogeny and evolutionary history of  
the silkworm. *Sci. China Life Sci.* 55:483-96. doi: 10.1007/s11427-012-4334-7.
- 410

- 412 Vanderstraete, M., Gouignard, N., Ahier, A., Morel, M., Vicogne, J., Dissous, C. 2013. The venus kinase  
receptor (VKR) family: structure and evolution. *BMC Genomics* 14:361. doi:10.1186/1471-2164-14-  
361.
- 414 Vanderstraete, M., Gouignard, N., Cailliau, K., Morel, M., Hahnel, S., Leutner, S., Beckmann, S.,  
416 Grevelding, C.G., Dissous, C. 2014. Venus kinase receptors control reproduction in the platyhelminth  
parasite *Schistosoma mansoni*. *PLoS Pathog.* 10:e1004138. doi: 10.1371/journal.ppat.1004138.
- 418 Vanhems, E., Delbos, M., Girardie, J. 1990. Insulin and neuroparsin promote neurite outgrowth in  
420 cultured locust CNS. *Eur. J. Neurosci.* 2:776-782. doi:10.1111/j.1460-9568.1990.tb00468.x.
- 422 Veenstra, J.A. 2010. What the loss of the hormone neuroparsin in the *melanogaster* subgroup of  
*Drosophila* can tell us about its function. *Insect Biochem. Mol. Biol.* 40:354-361. doi:  
424 10.1016/j.ibmb.2010.03.001.
- 426 Veenstra, J.A. 2014. The contribution of the genomes of a termite and a locust to our understanding of  
insect neuropeptides and neurohormones. *Front. Physiol.* 5:454. doi: 10.3389/fphys.2014.00454.
- 428 Veenstra, J.A. 2016. Similarities between decapod and insect neuropeptidomes. *PeerJ.* 4:e2043. doi:  
430 10.7717/peerj.2043.
- 432 Veenstra, J.A. 2019. Two Lys-vasopressin-like peptides, EFLamide, and other phasmid neuropeptides.  
*Gen. Comp. Endocrinol.* 278:3-11. doi: 10.1016/j.ygcen.2018.04.027.
- 434 Vogel, K.J., Brown, M.R., Strand, M.R. 2015. Ovary ecdysteroidogenic hormone requires a receptor  
436 tyrosine kinase to activate egg formation in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA.*  
112:5057-5062. doi: 10.1073/pnas.1501814112.
- 438 Xia, Q., Guo, Y., Zhang, Z., Li, D., Xuan, Z., Li, Z., Dai, F., Li, Y., Cheng, D., Li, R., Cheng, T., Jiang,  
440 T., Becquet, C., Xu, X., Liu, C., Zha, X., Fan, W., Lin, Y., Shen, Y., Jiang, L., Jensen, J., Hellmann, I.,  
Tang, S., Zhao, P., Xu, H., Yu, C., Zhang G., Li J., Cao, J., Liu, S., He, N., Zhou, Y., Liu, H., Zhao, J.,  
442 Ye, C., Du, Z., Pan, G., Zhao, A., Shao, H., Zeng, W., Wu, P., Li, C., Pan, M., Li, J., Yin, X., Li, D.,  
Wang, J., Zheng, H., Wang, W., Zhang, X., Li, S., Yang, H., Lu, C., Nielsen, R., Zhou, Z., Wang, J.,  
444 Xiang, Z., Wang, J. 2009. Complete resequencing of 40 genomes reveals domestication events and  
genes in silkworm (*Bombyx*). *Science.* 326:433-436. doi: 10.1126/science.1176620.
- 446 Xiang, H., Liu, X., Li, M., Zhu, Y., Wang, L., Cui, Y., Liu, L., Fang, G., Qian, H., Xu, A., Wang, W.,  
448 Zhan, S. 2018. The evolutionary road from wild moth to domestic silkworm. *Nat. Ecol. Evol.* 2:1268-  
1279. doi: 10.1038/s41559-018-0593-4.

## 450 **Legends to the figures**

452 **Fig. 1.** Alignment of selected Arthropod neuroparsin precursor sequences. Sequences are from  
454 *Tischeria quercitella* (GENO01134828.1), *Eumeta japonica* (BGZK01000877.1;  
join[241322..241488,242249..242387,243576..243797]), *Galleria mellonella* (XM\_026905025),



456 *Bombyx mandarina* (XM\_028184972.1), *Aedes aegypti* (U69542.1), *Tribolium castaneum*  
(XM\_015980588), *Apis mellifera* (XM\_026440542), *Locusta migratoria* (Y17710.1), *Procambarus*  
458 *clarkii* (GARH01002923.1) and *Carcinus maenas* (GFYV01020843.1). Note the very limited sequence  
similarity of the Lepidopteran neuroparsins with those from other Arthropods. Primary sequence  
460 similarity is so poor that Clustal Omega no longer aligns the cysteine residues, which are highlighted in  
red; other conserved amino acid residues are highlighted in black.

462

**Fig. 2.** Alignment of selected Lepidoptera neuroparsin precursor sequences. Sequences are from  
464 *Spodoptera litura* (GBBY01013502.1), *Helicoverpa armigera* (GFWI01343101.1), *Pieris rapae*  
(GGJY01013372.1), *Danaus plexippus* (OWR50203.1), *Papilio machaon* (XP\_014372220.1), *Galleria*  
466 *mellonella* (XM\_026905025), *Manduca sexta* (GETI01101303.1) and *Bombyx mandarina*  
(XM\_028184972.1). For more sequences see Table S1. Cysteine residues are highlighted in red and  
468 other conserved amino acid residues in black.

470 **Fig. 3.** *In situ* hybridization of neuroparsin mRNA in the brain of *Galleria mellonella* larva (**a**), and  
adult (**b**) with an amplification of the adult *pars intercerebralis* in (**c**). Note that these are whole mounts  
472 and consequently it is difficult to get all cells in focus in a single picture. Scale bars are 500  $\mu\text{m}$  in (**a**)  
and (**b**) and 200  $\mu\text{m}$  in (**c**).

474

**Fig. 4.** *Bombyx* neuroparsin genes. **a.** Comparative structure of the *Bombyx* neuroparsin genes as  
476 present in the genome assemblies. Green and purple indicate exons, green the untranslated and purple  
the translated parts. The numbers in the purple boxes and below the introns indicate their sizes in base  
478 pairs. Both genes have four exons, of which first is untranslated. TATA indicates the location of the  
TATA box. Note the large difference in the size of the intron between the first and second coding exon.  
480 Two different possible translations of the *B. mori* mRNA are represented, that would yield very  
different translated proteins. **b.** Direct comparison of the second exon (capitals) and the sequences  
482 immediately surrounding them (lower case). The start codon is indicated by an arrow. Note that apart  
from two SNPs, one before the start codon and one that causes a ile-val change, there is a 16 nucleotide  
484 deletion in the *B. mori* sequence. **c.** The possible translations of the two genes. In *B. mandarina* a  
conventional neuroparsin precursor is predicted. In *B. mori*, one predicts a very short peptide without  
486 any sequence similarity to neuroparsin but with a signal peptide. In case the first three start codons

488 were to be ignored, a protein could be produced that has a neuroparsin sequence, but lacks that a signal peptide and hence can not be expected to be secreted. Black highlighting indicates identical amino acid residues.



490 Figure 1.

```

Tischeria MW--TVI-AMVLIAGACSGHYGPSYKLPFSLT-----IPE
Eumeta MQLLRIM-VICAAVGYSAAYYHDSPOGLGLRDVIEPRDVIENYETAVQKRDE
Galleria MHPVLYV---MTTIGVCLAHYRG-----GYDARNLIEREE-----PQ
B.mandarina MPHIVFI-SLLIAAGTCFARYNG-----MYDSRNLIERDE-----PE
Aedes MSAGPVI-LLSVLLSGYIYWSLAQ-----
Tribolium MCPFHNFITIIILVLTVTVIIFSD-----
Apis MFAIQTIRIAVLLAI--VFLFD-----
Locusta MKATAAIVAAATLLAVTLFHRAE-----
Carcinus MASSCCRTATLILVCSCLLLLLQ-----
Procambarus MRRSSLITTT--VLVVVVVIFLQ-----

Tischeria KLVEPKLVECPVCRLESSIECEPF---DSVEGRKMLKKCRHGTFPNTMCNN
Eumeta YVLQLELVA CPVCDQF-----LDDECTVLKNGKRYKKCAFCTYINEVCGN
Galleria DDIEGEIIVECPVIGSSSNDWLP SNELCPIVRGNKRYKRCQRGTYVNTVCDN
B.mandarina INGDMEIVECPVED--SGHWLTSNEVCTIVKGHRKYKKCDHGTYINEVCGN
Aedes TNVLEIR--CKL-----YSGPAVQNTGECVHCAE-LNPCGK
Tribolium KGTAMIHLPCKRC-AT-----IQECNADPPQLCVFGEN-RDYCNR
Apis KSGYPS--IRQGTTS-----YCLGCGDSCHKCKYGIAMSSACGI
Locusta RNPISRS--CE-----GANCVVDLTRCEYCDV-TDFFCGR
Carcinus EASGAPR--CD-----RHDEE-APKNCKYGT-ODWCKN
Procambarus GTAGAPR--CS-----RQGGRAPPHNCKYGTV-SDWCGN

Tischeria KTD CYRGPGES CNEDE----EGIHCVPSLW CANNYVCKGTSDTGAATKTHSLD
Eumeta QRD CYRGPQYCTEKMAEDKLGAKCAHGYTCNAQLHECVG---LFNELDSLQ
Galleria RLD CYRGPKEQCTEKMDFDIYGQKCAPGYCCKNYLGVCTG---LEYNVESKT
B.mandarina RRD CFRGPGEQCTEKMEFDTYGQKCAPGYCCKNPKMGIQCG---MDYVLD SNE
Aedes -LSCLKGVGDKCGESTAGIIMSGK CASGLMCCG--GQCVGCKNGI--CDHRL
Tribolium -RVCSKGPGEKCGDRF---NILGTGEGELWCSNKDNRCHGCYIPTMAC---
Apis -VQCAKGPDEL CGGPQ---NYLGI CAEGMQCS C--NKCIGCSAEKFECSKTS
Locusta -KVCAKGPDKCGGPY---ELHGKCGVGMDCRC--GLCSGCSLHNLQCFFFE
Carcinus -GVCAKGPGETCGGYR---WSEKCGEGTFCS C--GICGCSFPD GKCGP--
Procambarus -PVCAKGPGETCGGEW---GENGACGAGTYCS C--GFCSGCSANL-ECWF--

Tischeria TFLYPVTWRNGFQ
Eumeta LAGLRHYIILPEDNTNSYIDKYRLARSHHLY
Galleria QWLLNPNRRYPLRAKRDV
B.mandarina LFLNPNHRRYSLQ RNTLKGPMSSHSLDAIRN
Aedes CP PRLTMNHHPFGLGLMAGSPQQQQPVGVFVPSLYKMFDDYSSSED
Tribolium ----YPDD
Apis NPCLPHRNSDRDLNIRKWSIAIN
Locusta GGLPSS-C
Carcinus ----TSIC
Procambarus ----GNFC

```

492 Figure 2.

```

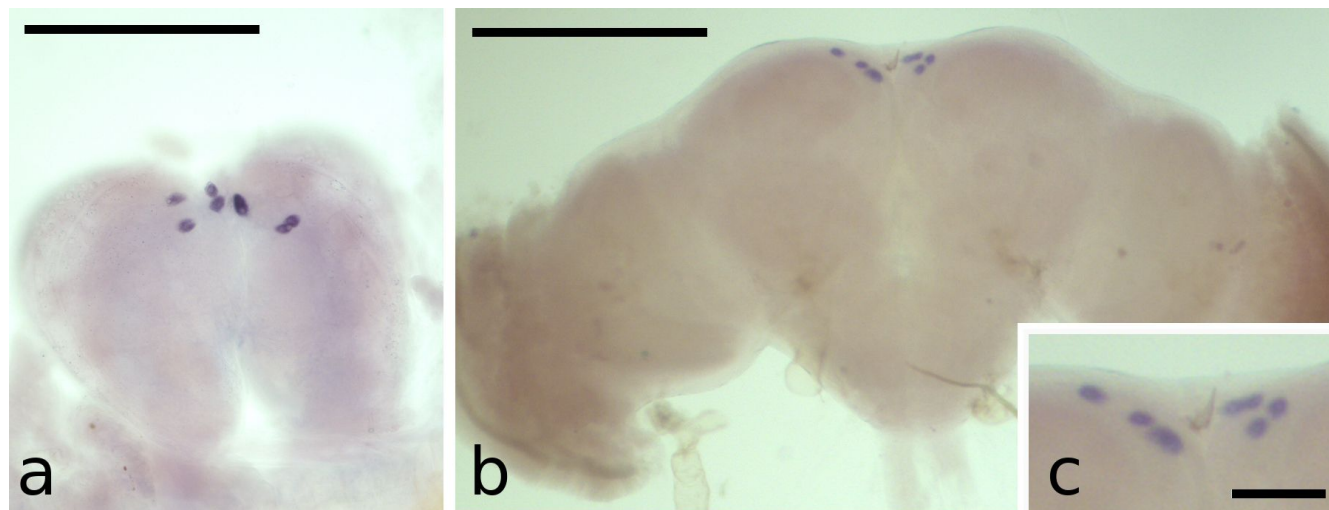
Spodoptera  MOSLVYLMML-VAFSSCOAHYRGSYDSRNFITREEP--GVLIECPVCDQ--TNHWM--
Helicoverpa MHTLVFLML-VTYSSCLAHYRGSYESRNYITREEP--K--GVLIECPVCPSP--E-GWD--
Pieris      MSPILYLL--VSAGCMAHYR-GYQGRNLIKEEHP-NEGILVECPVCYN--APWQFN
Danaus     MHLSIYILAS-VFYLTGHR-GYESRNLIVREEPFDL-EG-LVECPVCOY--QVSWL--
Papilio    MHLLAYILFFAATSCCLAHYRGGYESHRMFERNELHEDVDGVLVECPVCE--NGVL--
Galleria   MHPVLYVML--TIGVCLAHYRGGYDARNLIEREEPQDDIEGETVECPVICGSSNDWL--
Manduca    MOSLVYLMML-IIAGGVLAHYRGMVDSRNLIEREERMFDDGVEIVECPVVD--SGHWL--
B.mandarina MHPIVFLSLLLAAGTCFARYNGMYDSRNLIERDEPEINGDMEIVECPVCE--SGHWL--

Spodoptera  --ASTEDCPTRNGSKTFKICKFGTYINEVCGKRVDYSGPGEQCTEKREFDMYGKCKAPG
Helicoverpa --AIEEVC-IKECDKSYKLCRFGTYINTACNNRLDCLRGPPEKCTEKHVENRSNKKCTFG
Pieris      PDNGNPKCNITIGAORYKCKDRGSYFNEVCG-RRDCYRGPGEPCTEKMEDEYGAKCAF
Danaus     --PTREVCNITIGNKKYKRCRMGNYTNTVCGNRLDCFRGPGEQCTESMENDYQKCAR
Papilio    --PTHEHCNIIRGNKYYKKEFGTYKNEVCGNRRDCYRGPPEQCTEKMEIDVYGOKCAY
Galleria   --PSNELCPIVRGNKRYKRCORGTYNNTVCGNRLDCYRGPKEQCTEKMEFDIYGOKCAP
Manduca    --PSNEECTIRKGNKYYKCELGTYINEVCGNRRDCFRGAEEQCTEKMEVDIYGOKCAG
B.mandarina --TSNEVCTIVKGHRKYKCKDHGTYINEVCGNRRDCFRGPGEQCTEKMEFDIYGOKCAP

Spodoptera  YYCNKMEGVCTGLEFQLGNNWQWKYF---YTGVRSELRNQPEDSKFSFEGVRN
Helicoverpa YECNQKEGVCTGPENRPEIYWPYRFGL---YPDRSSSKSTAE---AFEDMTN
Pieris      YDCDFSYHVCTGIGFSISIHDLRSF--PRLYPLRPLNEGENQA--KSMILT
Danaus     YDCDGFHFVCTGYGYTINSHMRWLLNHVSRYPGNQONDEQLREK--ALYLA
Papilio    YYCDNLYNVCVGLDYSVERPFG-IYNHLYQYPLRDDKLIV
Galleria   YYCNKYLGVCCTGLEYNVESKTQWLLNPNRYPPLRAKRDV
Manduca    YYCNKLMGICOGLEYVPDSTQQFLLNPHHRYSLQORSGIKTRAA--NHFLDTIRN
B.mandarina YYCNKPMGICOGMDYVLDSENELFLLNPHHRYSLQORNTLKGPM--HSLLDAIRN

```

494 Figure 3.



496 Figure 4.

