

# The *Drosophila* gene CG9918 codes for a pyrokinin-1 receptor<sup>☆</sup>

Giuseppe Cazzamali, Malene Torp, Frank Hauser, Michael Williamson,  
Cornelis J.P. Grimmelikhuijzen<sup>\*</sup>

Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, Universitetsparken 15,  
DK-2100 Copenhagen, Denmark

Received 12 July 2005  
Available online 19 July 2005

## Abstract

The database from the *Drosophila* Genome Project contains a gene, CG9918, annotated to code for a G protein-coupled receptor. We cloned the cDNA of this gene and functionally expressed it in Chinese hamster ovary cells. We tested a library of about 25 *Drosophila* and other insect neuropeptides, and seven insect biogenic amines on the expressed receptor and found that it was activated by low concentrations of the *Drosophila* neuropeptide, pyrokinin-1 (TGPSASSGLWFGPRLamide; EC<sub>50</sub>,  $5 \times 10^{-8}$  M). The receptor was also activated by other *Drosophila* neuropeptides, terminating with the sequence PRLamide (Hug-γ, ecdysis-triggering-hormone-1, pyrokinin-2), but in these cases about six to eight times higher concentrations were needed. The receptor was not activated by *Drosophila* neuropeptides, containing a C-terminal PRLamide sequence (such as ecdysis-triggering-hormone-2), or PRVamide (such as capa-1 and -2), or other neuropeptides and biogenic amines not related to the pyrokinins. This paper is the first conclusive report that CG9918 is a *Drosophila* pyrokinin-1 receptor gene.

© 2005 Elsevier Inc. All rights reserved.

**Keywords:** G protein-coupled receptor; GPCR; Neurohormone; Neuropeptide; Peptide hormone; Neuromedin; Pyrokinin; Sex pheromone; Hugin gene; Capability gene; Diapause; Ecdysis; Evolution; Insect; Honey bee; Mosquito; Malaria

The presence of a sequenced genome of *Drosophila melanogaster* [1] gives us the opportunity to identify all proteins and, therefore, to understand all biochemical and physiological processes that occur in an insect. Our research group is especially interested in G protein-coupled receptors (GPCRs), because these proteins and their ligands (neuropeptides, protein hormones, and biogenic amines) occupy a high hierarchic position in the physiology of insects and steer important processes, such as reproduction, development, feeding, sugar and

water homeostasis, and behavior. Of the about 14,000 genes present in *Drosophila*, 47 genes code for neuropeptide and protein hormone GPCRs, and 20 for biogenic amine GPCRs [2]. About half of these genes have been deorphanized and characterized, so far, and it can be expected that the characterization of the remaining 50% of these GPCR genes, which might be somewhat more difficult to deorphanize, will occur within the next few years [2]. The completion of this neurohormone GPCR characterization project is important, because it will supply us with a whole new view on insect endocrinology and physiology.

*Drosophila* produces a family of seven neuropeptides that have the C-terminal sequence PRL/I/Vamide in common ([3]; Table 1). Two of them are defined as pyrokinins (Drm-PK-1 and -2), which are characterized by the C-terminal FXPRLamide sequence (Table 1) [4,5]. The pyrokinins are important and ubiquitous insect

<sup>☆</sup> The nucleotide sequences reported in this paper have been submitted to the GenBank/EBI Data Bank with Accession Nos. AF368273, BK005274, and DQ103706.

<sup>\*</sup> Corresponding author. Fax: +45 35321200.

E-mail address: [cgrimmelikhuijzen@bi.ku.dk](mailto:cgrimmelikhuijzen@bi.ku.dk) (C.J.P. Grimmelikhuijzen).

URL: <http://www.bi.ku.dk/staff/staff-vip-details.asp?ID=90> (C.J.P. Grimmelikhuijzen).

Table 1

Amino acid sequences of some structurally related insect neuropeptides and their potencies to activate the CG9918 receptor

Name	Structure	Species	EC <sub>50</sub> (M) with the CG9918 receptor
Drm-PK-1 (Drm-myotropin, capa-3)	TG <b>PSASSGLWFGPR</b> Lamide	<i>D. melanogaster</i>	$5 \times 10^{-8}$
Hug- $\gamma$	pQLQ <b>SN</b> GEPAYRV <b>RTPR</b> Lamide	<i>D. melanogaster</i>	$3 \times 10^{-7}$
Drm-PK-2	S <b>VPFKPR</b> Lamide	<i>D. melanogaster</i>	$4 \times 10^{-7}$
Drm-ETH-1	DDSSPG <b>FFLK</b> ITKN <b>VPRL</b> Lamide	<i>D. melanogaster</i>	$3 \times 10^{-7}$
Capa-1	GANMGLYA <b>FP</b> RVamide	<i>D. melanogaster</i>	NA
Capa-2	ASGLVA <b>FP</b> RVamide	<i>D. melanogaster</i>	NA
Drm-ETH-2	GENFAIK <b>NLKTIP</b> RIamide	<i>D. melanogaster</i>	NA
Leucopyrokinin (Lem-PK)	pQT <b>STFTPR</b> Lamide	<i>L. maderae</i>	$>1 \times 10^{-6}$

NA, not active in concentrations up to  $10^{-5}$  M. *L. maderae*, *Leucophea maderae* (cockroach).Bold highlights the pyrokinin consensus sequence (FXPR<sub>L</sub>amide) or its C-terminal PRLamide portion.

neuropeptides that play a central role in diverse physiological processes, such as insect sex pheromone production, diapause, pupariation, and gut motility [4–8]. The seven *Drosophila* neuropeptides from Table 1 are produced by three preprohormones [3]. One contains capa-1, capa-2, and Drm-PK-1, one contains Hug- $\gamma$  and Drm-PK-2, and one contains ecdysis-triggering-hormone (ETH)-1 and -2 [3]. We have previously cloned and characterized the receptors for most of the neuropeptides mentioned in Table 1. The two ETH receptors (the two splicing variants from gene CG5911) are specific for ETH-1 and -2, and do not cross-react with the other peptides from Table 1 [9]. Also, the capa receptor (coded for by the gene CG14575) is specific for capa-1 and -2 [10]. In contrast, the pyrokinin receptors appear to be more promiscuous. We have cloned and characterized two *Drosophila* pyrokinin-2 receptors (encoded by genes CG8784 and CG8795) that get activated by low concentrations of Drm-PK-2 (EC<sub>50</sub>'s below  $10^{-9}$  M) [11]. However, also Hug- $\gamma$  and ETH-1 could activate these two receptors, although 14–30 (Hug- $\gamma$ ) and 40–200 (ETH-1) times higher concentrations were needed [11]. Drm-PK-1, albeit a pyrokinin, did hardly activate the two pyrokinin-2 receptors [11].

The existence of a Drm-PK-1 receptor has not been convincingly demonstrated, so far. Park et al. [12] have cloned a *Drosophila* gene (CG9918) and expressed it in *Xenopus* oocytes. Only very high, non-physiological concentrations (above  $10^{-5}$  M) of Drm-PK-1 did activate the receptor and, because of the high concentrations of peptide needed, an EC<sub>50</sub> could not be determined [12]. In the present paper, we will report on the successful identification of the *Drosophila* Drm-PK-1 receptor.

## Materials and methods

Database screening was carried out, using the Berkeley *Drosophila* Genome Project BLAST server, and genomic DNA sequences were analyzed for complete gene structures, using the Genscan Web Server at the Massachusetts Institute of Technology. cDNA from *D. melanogaster* third instar larvae (Canton S) was used as template. As primers for our initial PCR experiments, we used the sense primer

5'-ATACCCGTAACGGTAGTCTAC-3' (corresponding to nucleotide positions 61–81 of Fig. 1) and the antisense primer 5'-AGCTGG AATGAGTGCTTCACT-3' (corresponding to nucleotide positions 531–551 of Fig. 1). The PCR program was 1 cycle of 95 °C for 30 s, 58 °C for 30 s, 68 °C for 1 min followed by 30 cycles of 95 °C for 30 s, 58 °C for 30 s, 68 °C for 1 min, and a final extension step of 68 °C for 10 min. PCR was carried out using the Advantage2 PCR enzyme system (Clontech). The SMART RACE cDNA kit (Clontech) was used for the rapid amplification of cDNA end (RACE) reactions. The 3'-RACE reactions were made with the sense primer 5'-CTGAAGC ACTCATTCCAGCTGTCCACGTTTC-3' (corresponding to nucleotide positions 532–562 of Fig. 1). The 5'-RACE reactions were carried out with the antisense primer, 5'-GAGTAGACTACCGTTACGGGT ATCACGATGG-3' (corresponding to nucleotide positions 53–83 of Fig. 1). All PCR products were cloned into pCR4-TOPO (Invitrogen), using the TOPO TA Cloning method (Invitrogen), and sequenced.

Chinese hamster ovary (CHO) cells were grown as described previously [13]. To amplify a full-length cDNA coding for the receptor, the following primers were applied: the sense primer 5'-CAAGCTTAA GATGTCCGCTGGCAAT-3' (corresponding to nucleotide positions 1–15 of Fig. 1) and the antisense primer 5'-TCTAGATTAGTTGAC TTGGACACCGATCATG-3' (corresponding to nucleotide positions 1269–1293 of Fig. 1). The *Xba*I and *Hind*III restriction sites that had been incorporated into the above primers facilitated the subcloning into the pcDNA3 vector (Invitrogen). The insert was fully sequenced and the plasmid was transfected into CHO cells, using the method previously described [13]. The bioluminescence assay was described earlier [13,14].

DNA sequence compilation, and nucleotide and amino acid sequence comparisons were performed using DNASTAR [11].

## Results

Already in 1999, one year before the completion of the *Drosophila* Genome Project [1], we cloned a gene, coding for a GPCR, which was later assigned the CG No. CG9918. Fig. 1 shows the cDNA of this gene. It is 1617 nucleotides long, contains a polyadenylation site in its untranslated 3'-region and an in-frame stop codon, preceding the start codon, in its untranslated 5'-region. The coding region is 430 amino acid residues long and contains all the characteristics of a GPCR, including seven transmembrane  $\alpha$ -helices and an ERY (Glu-Arg-Tyr) consensus sequence shortly after TM III, showing that it belongs to the rhodopsin-like GPCR family (family-1 or family-A) [15].

ACGCGGATAGTTGAGGGACATCGAGTCGACGCCG																										-1
AGCCGCAGCAGTTAAACGGCAGTGGAAATTTAGCGCTGTTAGTGGAAATTTGTGAAAACAAACCGGACCTTAAAGGGGTTCCAGGCCGCCGCGGCC																										-1
TM I																										
ATG	TCC	GCT	GGC	AAT	ATG	AGC	CAT	GAT	CTT	GGA	CCG	CCT	CGC	GAT	CCG	CTG	GCC	ATC	GTG	ATA	CCC	GTA	ACG	GTA	75	
Met	Ser	Ala	Gly	Asn	Met	Ser	His	Asp	Leu	Gly	Pro	Pro	Arg	Asp	Pro	Leu	Ala	Ile	Val	Ile	Pro	Val	Thr	Val	25	
▲																										
GTC	TAC	TCC	CTG	ATT	TTC	ATA	ACC	GGT	GTG	GTT	GGC	AAC	ATA	AGT	ACC	TGC	ATT	GTG	ATT	AAG	AAG	AAC	CGT	TCA	150	
Val	Tyr	Ser	Leu	Ile	Phe	Ile	Thr	Gly	Val	Val	Gly	Asn	Ile	Ser	Thr	Cys	Ile	Val	Ile	Lys	Lys	Asn	Arg	Ser	50	
TM II																										
ATG	CAC	ACG	GCC	ACG	AAT	TAC	TAC	CTC	TTT	TCG	CTG	GCC	ATC	TCG	GAT	TTC	CTG	CTC	CTG	TTG	TCG	GGC	GTT	CCG	225	
Met	His	Thr	Ala	Thr	Asn	Tyr	Tyr	Leu	Phe	Ser	Leu	Ala	Ile	Ser	Asp	Phe	Leu	Leu	Leu	Leu	Ser	Gly	Val	Pro	75	
TM III																										
CAG	GAG	GTG	TCC	TAC	ATC	TGG	TCC	AAG	TAC	CCG	TAC	GTG	TTT	GGG	GAG	TAC	ATC	TGC	ATC	GGA	CGT	GGT	CTG	TTG	300	
Gln	Glu	Val	Ser	Tyr	Ile	Trp	Ser	Lys	Tyr	Pro	Tyr	Val	Phe	Gly	Glu	Tyr	Ile	Cys	Ile	Gly	Arg	Gly	Leu	Leu	100	
TM IV																										
GCG	GAG	ACA	TCG	GCG	AAT	GCC	ACG	GTG	CTA	ACG	ATT	ACG	GCC	TTC	ACG	GTG	GAG	CGG	TAT	ATT	GCC	ATT	TGC	CAT	375	
Ala	Glu	Thr	Ser	Ala	Asn	Ala	Thr	Val	Leu	Thr	Ile	Thr	Ala	Phe	Thr	Val	Glu	Arg	Tyr	Ile	Ala	Ile	Cys	His	125	
TM V																										
CCG	TTT	CTG	GGC	CAG	GCC	ATG	AGT	AAA	CTC	AGT	CGC	GCC	ATT	CGC	ATC	ATC	GTC	CTG	GTT	TGG	ATT	ATG	GCC	ATA	450	
Pro	Phe	Leu	Gly	Gln	Ala	Met	Ser	Lys	Leu	Ser	Arg	Ala	Ile	Arg	Ile	Ile	Val	Leu	Val	Trp	Ile	Met	Ala	Ile	150	
TM VI																										
GTT	ACG	GCC	ATT	CCG	CAG	GCT	GCC	CAA	TTT	GGA	ATC	GAG	CAC	TAT	TCG	GGA	GTG	GAG	CAA	TGC	GGC	ATA	GTG	CGG	525	
Val	Thr	Ala	Ile	Pro	Gln	Ala	Ala	Gln	Phe	Gly	Ile	Glu	His	Tyr	Ser	Gly	Val	Glu	Gln	Cys	Gly	Ile	Val	Arg	175	
TM VII																										
GTC	ATA	GTG	AAG	CAC	TCA	TTC	CAG	CTG	TCC	ACG	TTC	ATA	TTC	TTC	CTG	GCT	CCG	ATG	TCT	ATT	ATC	CTA	GTG	TTG	600	
Val	Ile	Val	Lys	His	Ser	Phe	Gln	Leu	Ser	Thr	Phe	Ile	Phe	Phe	Leu	Ala	Pro	Met	Ser	Ile	Ile	Leu	Val	Leu	200	
TM VIII																										
TAC	CTA	CTT	ATC	GGT	GTG	CAC	CTG	TAT	CGA	TCC	ACT	TTG	GTG	GAG	GGT	CCT	GCC	TCG	GTT	GCC	AGA	CGG	CAG	CAG	675	
Tyr	Leu	Leu	Ile	Gly	Val	His	Leu	Tyr	Arg	Ser	Thr	Leu	Val	Glu	Gly	Pro	Ala	Ser	Val	Ala	Arg	Arg	Gln	Gln	225	
TM IX																										
CTG	AAG	AGT	GTG	CCC	AGT	GAT	ACG	ATC	CTA	TAT	CGC	TAT	GGT	GGA	TCC	GGT	ACC	GCT	ATG	AGT	TTC	AAC	GGC	GGA	750	
Leu	Lys	Ser	Val	Pro	Ser	Asp	Thr	Ile	Leu	Tyr	Arg	Tyr	Gly	Gly	Ser	Gly	Thr	Ala	Met	Ser	Phe	Asn	Gly	Gly	250	
TM X																										
GGA	AGT	GGA	GCA	GGG	ACA	GCG	GGC	TTG	ATG	GGC	GGC	TCG	GGG	GCT	CAA	CTT	AGC	TCG	GTG	AGA	GGT	CGG	CTC	AAT	825	
Gly	Ser	Gly	Ala	Gly	Thr	Ala	Gly	Leu	Met	Gly	Gly	Ser	Gly	Ala	Gln	Leu	Ser	Ser	Val	Arg	Gly	Arg	Leu	Asn	275	
↓1																										
CAC	TAT	GGC	ACC	CGG	CGA	GTA	CTC	AGG	ATG	CTA	GTG	GCC	GTG	GTG	GTG	TGC	TTC	TTC	CTC	TGC	TAG	GCC	CCC	TTC	900	
His	Tyr	Gly	Thr	Arg	Arg	Val	Leu	Arg	Met	Leu	Val	Ala	Val	Val	Val	Cys	Phe	Phe	Leu	Cys	Trp	Ala	Pro	Phe	300	
TM XI																										
CAC	GCC	CAG	CGA	CTG	ATT	GCC	ATC	TAC	GCC	CCT	GCA	CGG	GGG	GCC	AAA	CTG	CGG	GAT	CAG	CAC	GAG	TTT	GTC	TAC	975	
His	Ala	Gln	Arg	Leu	Ile	Ala	Ile	Tyr	Ala	Pro	Ala	Arg	Gly	Ala	Lys	Leu	Arg	Asp	Gln	His	Glu	Phe	Val	Tyr	325	
TM XII																										
ACG	GTG	ATG	ACC	TAT	GTC	TCC	GGT	GTC	CTC	TAC	TAT	TTG	TCC	ACC	TGC	ATC	AAC	CCG	CTG	TTA	TAT	AAC	ATT	ATG	1050	
Thr	Val	Met	Thr	Tyr	Val	Ser	Gly	Val	Leu	Tyr	Tyr	Leu	Ser	Thr	Cys	Ile	Asn	Pro	Leu	Leu	Tyr	Asn	Ile	Met	350	
↓2																										
AGC	CAC	AAG	TTC	CGA	GAG	GCA	TTC	AAG	GCC	GTT	CTG	TTT	GGC	AAG	AAG	GTA	TCG	AAG	GGT	TCG	CTG	AAT	TCG	CGA	1125	
Ser	His	Lys	Phe	Arg	Glu	Ala	Phe	Lys	Gly	Ala	Val	Leu	Phe	Gly	Lys	Lys	Val	Ser	Lys	Gly	Ser	Leu	Asn	Ser	Arg	375
TM XIII																										
AAC	AAC	ATC	GAA	TCG	CGC	CGC	CTG	AGG	AGG	GCA	CTA	ACC	AAT	TCC	AGT	CAA	ACG	CAG	CGC	TTC	TCC	ATT	GAG	TCG	1200	
Asn	Asn	Ile	Glu	Ser	Arg	Arg	Leu	Arg	Arg	Ala	Leu	Thr	Asn	Ser	Ser	Gln	Thr	Gln	Arg	Phe	Ser	Ile	Glu	Ser	400	
↓3																										
GCG	GAG	CAG	CCG	AAA	CCG	TCG	ATA	ATG	CAG	AAT	CCG	ACG	AAC	AAG	CCG	CCA	GTC	GCT	CAG	TAC	GCC	ATG	ATC	1275		
Ala	Glu	Gln	Pro	Lys	Pro	Ser	Ile	Met	Gln	Asn	Pro	Thr	Asn	Lys	Pro	Pro	Val	Ala	Ala	Gln	Tyr	Ala	Met	Ile	425	
TM XIV																										
GGT	GTC	CAA	GTC	AAC	TAA	CTGATAAACTCGAACTCACTCTCTCGTGC	GGCTTCCTGATGCTAATTTATTTTGAATTCTGTTTAAGCCAGTTATT	1368																		
Gly	Val	Gln	Val	Asn	*																	430				
TM XV																										
TGTCGTCTATTGTCTATTGCTGCTACCAAAAGCGTTTTGTTTAAACTGCTTTTGCATTTCATTTGCATTGGCAATTAAGTTAATCACAGTCGAGCG																										1467
CACACTCTCCCTC(A) <sub>n</sub>																										1481

Fig. 1. cDNA and deduced amino acid sequence of the transcript (GenBank Accession No. [AF368273](#)) from gene CG9918. Nucleotides are numbered from 5'- to 3'-end and the amino acid residues are numbered starting with the first ATG codon in the open reading frame. The introns are indicated by arrows (numbered 1–3) and the exon nucleotides, bordering these introns, are highlighted in grey. The seven transmembrane  $\alpha$ -helices are boxed and labelled TM I–VII. The in-frame stop codon in the 5'-noncoding region is underlined. The translation termination codon is indicated by an asterisk. The putative polyadenylation signal in the 3'-noncoding region is underlined twice. The putative N-glycosylation site in the extracellular N terminus is indicated by a triangle.

A comparison of the cDNA of Fig. 1 with the genomic sequence of CG9918 revealed only a few nucleotide differences (Table 2). These differences did not lead to a change in amino acid residues in the receptor protein and probably represent allelic variations of the gene. The comparison of cDNA and genomic sequences also revealed that the gene contained four exons and three introns (Table 3).

We stably expressed the coding region of CG9918 in CHO cells that also were stably expressing the  $\alpha$  subunit of the promiscuous G protein G-16 [14]. These cells were also transiently transfected with DNA, coding for the protein apoaeguorin. Three hours before the assay, coelenterazine was added to the cells. Activation of the expressed receptors and G-16 in these pretreated CHO cells would result in an  $\text{IP}_3/\text{Ca}^{2+}$ -mediated biolumines-

Table 2

Nucleotide differences between the receptor cDNA of Fig. 1 and the corresponding genomic sequences from the Berkeley “*Drosophila* Genome Project”

Position of the nucleotide in the cDNA	Type of nucleotide in the gene	Type of nucleotide in the cDNA	Change in amino acid
39	A	T	—
75	G	A	—
87	C	G	—
105	A	G	—
108	G	T	—

Table 3

Intron/exon boundaries of the CG9918 receptor gene

Intron	5'-donor	Intron size (bp)	3'-acceptor	Intron phase
1	TAG gtaatta... Val	528	... ccttttag TGG Val	1
2	AAG gtgagta... Lys	133	... ccttttag GCC Ala	3
3	CAG gcaagtg... Gln	89	... catccag AAT Asn	3

cence response that could easily be measured and quantified [13,14].

We tested a neuropeptide library of 25 *Drosophila* and other insect neuropeptides, and seven insect biogenic amines on the transfected CHO cells. Low concentrations of Drm-PK-1 gave a strong bioluminescence response in CHO/G-16 cells transfected with the receptor DNA (Fig. 2B), whereas non-transfected cells did not react (Fig. 2A). A dose–response curve of the bioluminescence response yielded an  $EC_{50}$  of  $5 \times 10^{-8}$  M (Fig. 2C).

There are other *Drosophila* neuropeptides that are structurally related to Drm-PK-1 (Table 1). From these,

only Hug- $\gamma$ , Drm-PK-2, and Drm-ETH-1 could activate the receptor, but six to eight times higher concentrations were needed to elicit the same effects as Drm-PK-1 (Table 1). All other 20 insect neuropeptides tested and the seven insect biogenic amines (listed in [11]) did not activate the receptor, except for leucopyrokinin, which is a pyrokinin isolated from the cockroach, *Leucophaea maderae* [4] (Table 1).

The Drm-PK-1 receptor (CG9918) is structurally closely related to the two Drm-PK-2 receptors (CG8795 and CG8784) (Fig. 3A) [11]. There is 46% sequence identity between CG9918 and CG8795 (60% similarity), whereas these numbers for CG9918 and CG8784 are 48% (60%). A blast search of the genomic databases from the malaria mosquito *Anopheles gambiae* and the honey bee *Apis mellifera* revealed an orthologue to the Drm-PK-1 receptor gene in each of these two model insects (Ang-PR-1 and Ame-PR-1, Fig. 3A). All five receptor genes have two introns in common (with the same intron phasings), supporting the view that these receptors are evolutionarily closely related. A phylogenetic tree analysis of the five receptors from Fig. 3A further confirms their close evolutionary relationships (Fig. 3B). The tree also shows that the mosquito and honey bee receptors are orthologues to the Drm-PK-1 receptor (CG9918) and not to the Drm-PK-2 receptors (CG8795, CG8784).

## Discussion

Already in 1999, we cloned the cDNA of a GPCR that later, by the *Drosophila* Genome Project, was assigned the CG No. CG9918. It has long been unclear, what the intrinsic ligand of the CG9918 receptor was.

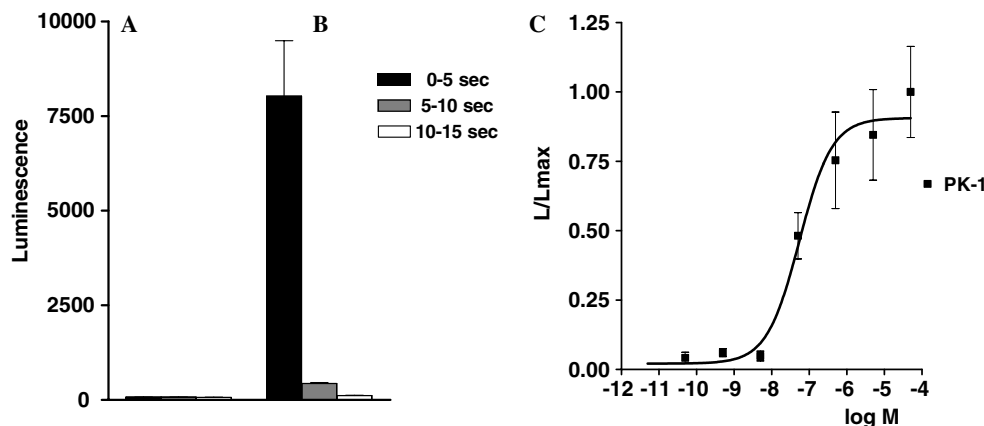


Fig. 2. Bioluminescence responses of non-transfected CHO/G-16 cells (A) and of CHO/G-16 cells, expressing the CG9918 gene (B) 0–5 s (black), 5–10 s (grey), and 10–15 s (white) after addition of  $5 \times 10^{-7}$  M Drm-PK-1. (C) Dose–response curve of the effect of Drm-PK-1 on CHO/G-16/CG9918 cells ( $EC_{50}$ ,  $5 \times 10^{-8}$  M). In all panels, the SEMs are given as vertical bars, which are sometimes smaller than the symbols used (squares or lines). In these cases only the symbols are given. In addition to Drm-PK-1, the CG9918 receptor is also activated by Hug- $\gamma$ , Drm-PK-2, Drm-ETH-1, and Lem-PK (Table 1), although much higher concentrations are needed. Twenty other insect neuropeptides (listed in Table 1 and [11]) and 7 biogenic amines (listed in [11]) did not activate the receptor (tested up to  $10^{-5}$  M).

<b>A</b>	Ang-PR-1	-----MSTSLKNN	8
	Ame-PR-1	-----MDHLTTILPNV	11
	CG8795	-----MLPTNSSGVLATDQLFHNKFLN	25
	CG8784	MLQGVAITTANDSNDDGLNQSFMAHVSPSPNQSPSIVGVGIGASSTMANPSESPEMLLLKNDKFLTHVAH	70
	CG9918	-----MSAGNMSHDLGPPRDLAIVIPVTVVYSLIFITGVVGNISTCIVIKNRSMT	53
	Ang-PR-1	PYELLEILTDDNESILSDGVESLTEMYGPKRDPLYVVIPIITIIYLLIFITGVVGNISTCIVIARNRSMT	78
	Ame-PR-1	SNLSFLHSQNFSLYTNHTNEINSPSSVPHRDSLHIVIPVTIIYVSIFVTGIIGNISTCIVIARNKSMHT	81
	CG8795	LTQVLNISADNLTSLQGLEPEELLPTVIPTPLSLLATLSVGYALIFVAGVLGNLITCIVISRNNFMHT	95
	CG8784	LLNITTENLSNLLGSTNGTNASTMAADSEVDESILTRTALTVCYALIFVAGVLGNLITCIVISRNNFMHT	140
	CG9918	ATNYLFLSLAISDFLLLSGVPOEVSYSKYYPVFGVEYICIGRGLLAETSANATVLTITAFTVERYIAI	123
	Ang-PR-1	ATNYLFLSLAVSDFLLLSGVPOEYIYFWSKYYPVFGETFCVLRGIAAEMSANATVLTITAFTIERYFAI	148
	Ame-PR-1	ATNYLFLSLAVSDFLLLSGLPAEIMYVWCKYPVFGEGFCILRGLAETSANATVLTITAFTVERYIAI	151
	CG8795	ATNFYLFNLAISDMILLCSGMPQDLYNLWHPDNPFSDSICILESVLSETAANATVLTITAFTVERYIAI	165
	CG8784	ATNFYLFNLAVSDFLLLSGVIPQELYNLYWPDMPFTDAMCIMGSVLSEMAANATVLTITAFTVERYIAI	210
	CG9918	CHPFLGQAMSKLSRAIRITVLVWIMAVTAIPQAAQFGIEHYSG---VEQCIVRVIVKHSFOLSTFIFF	190
	Ang-PR-1	CHPFLSHTMSKLSRAVRFICVILIAIVSAIPQALQFGVTNQGG---IDQCVVKRIIIQHSFELSTFLFF	215
	Ame-PR-1	CHPFLSQTMSKLTAVKILVVLVALSFALQALQFGLVQKHAPEVVMCTVKRIILQHSFELSTFLFF	221
	CG8795	CHPFRQHTMSKLSRAVKFIFAIWIAALLLALQAIQFSVVMQGM---GTSCMTKNDFFAHVFAVSGFLFF	232
	CG8784	CHPFRQHTMSKLSRAIKFIFAIWIAALLLALQAMQFSVVYQNE---GYSCTMENDFYAHVFAVSGFLFF	277
	CG9918	LAPMSIILVLYLLIGVHLYRSTLVEGPASVARRQQLKSVPSDTILRYGSGGTAMSFNGGSGAGTAGLM	260
	Ang-PR-1	FAPMTMITLYALIGLKLRTSTLMQORDTLQRR-----	248
	Ame-PR-1	VVPMSLITVLYALIGLKLKSNMMKRSRGREMG-----	254
	CG8795	GGPMTAICVLYVLIGVKLKRSRLQ---ALPRR-----	262
	CG8784	GGPMTAICVLYVLIGVKLKRSRLQ---SLPRR-----	307
	CG9918	GGSGAQLSSVRGRNLHYGTRVRVRLMVAVVVCFFLCWAPFHAQRLIAIYAPARGAKLRDQH---EFVYTV	327
	Ang-PR-1	---TQPSPRQS FANSQGSRRVLKMLVAVVVAFFICWAPFHAQRLVYIYGVNTNHQPSDPL--ILKLFII	312
	Ame-PR-1	---GSCRHQTGR-SSRRVLKMLVAVVIAFFICWAPFHVQRLIAIYGTNSEDHISSENSEWIEFLYLL	316
	CG8795	---CYDVNRGISAQTRVIRMLVAVAVAFFICWAPFHAQRLMAVYGS---TSGIBSQWFN-DYFSI	320
	CG8784	---TFDANRGLNAQGRVIRMLVAVAVAFFICWAPFHAQRLMAVYGLNLINIGISRDAFN-DYFRI	368
	CG9918	MTYVSGVLYYLSTCINPLLYNIMSHKFREAFKAVLFGKKVSKG-----SLNSRNIESRRL	383
	Ang-PR-1	TTYISGILYYLSTCINPLLYNIMSNKFRQAFK-----	344
	Ame-PR-1	MTYISGVFYIYSTINPILYNIMSNKFRVAFM-----	348
	CG8795	LDYTSGLVLYFLSTCINPLLYNIMSHKFREAFKVTILARHFGGLG---KNQGRGLPHTYSALRRNQTSRL	387
	CG8784	LDYTSGLVLYFLSTCINPLLYNIMSHKFREAFKILTTRQFGLARNHHHQSQHHQHNYSSALLR-QNGSMRL	437
	CG9918	RRA-----LTNSSQTRFSIESAEQ-----	403
	CG8795	HTTDSVRTTMTSMATTTTGLNGSANGSNGTTTGQSVRLNRVLSVQMCGQNRSRQDLFDNPRRLQTOQ	457
	CG8784	QPASCSVNNALEPYGSYRVVQFRCDANHQSLQDSIRTTTITTTINSNSMAAGNGVGGGAGGGRRRLR	507
	CG9918	-----PKPSIMQNPTNKPPVAAQYAMIGQVN-----	430
	CG8795	ISQLSSVGDAHSLLEEDLQFPGEPLQRQPTMCSIDELTDLAISRSRLKLTRITRPPGGVTGGVAGGSTT	527
	CG8784	KQELYPVPVGTAVPHRMLQAQVSQLSSLDANSLLAEVVDHRYASGRAKRALATKSGALLVTPPQSGD	577
	CG8795	VAAGSGGVSGDESSGKVRKAKVKVLKSSSPFKGLRTKFNWRARKGSHKPHEKGATVNGGDTEERAFF--	595
	CG8784	PSEVSQPATRLKLTRVISRRDEVANASTPPFCGSHSLDPETCQSASVAGRSSRKFPWRKRRQKTEDPSS	647
	CG8784	EGLTYGSPKQ-----	658

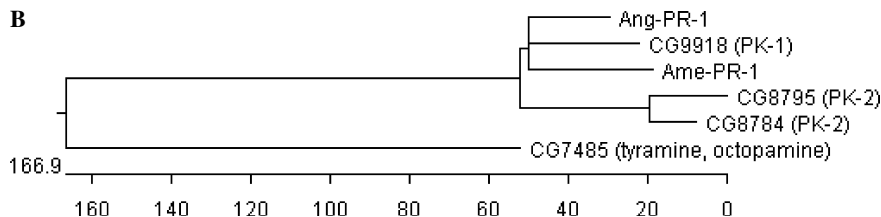


Fig. 3. (A) Amino acid sequence comparisons between the Drm-PK-1 receptor (encoded by CG9918), the orthologues from the malaria mosquito *A. gambiae* (Ang-PR-1, with GenBank Accession No. DQ103706) and the honey bee *A. mellifera* (Ame-PR-1, Accession No. BK005274), and the two *Drosophila* Drm-PK-2 receptors (encoded by genes CG8795, Accession No. AY277899; and CG8784, Accession No. AY277898). Amino acid residues that are identical between the CG9918 gene product and at least one of the other proteins are highlighted in grey. The seven transmembrane  $\alpha$ -helices are indicated by TM I–VII. The two common introns are indicated by vertical boxes. (B) A phylogenetic tree analysis of the five receptor proteins from (A). The *Drosophila* tyramine/octopamine receptor gene CG7484 [18] is used as an outgroup.

Park et al. [12], who expressed the cDNA in frog oocytes, found that the receptor could be activated by Drm-PK-1, but only very high, non-physiological concentrations (above  $10^{-5}$  M) could elicit responses and no  $EC_{50}$  value could be determined. Furthermore, Park

et al. [12] found that the receptor was insensitive to any of the Drm-PK-1-related neuropeptides given in Table 1.

We have now expressed the CG9918 cDNA in CHO/G-16 cells and established that the receptor was a Drm-PK-1 receptor with an  $EC_{50}$  for its ligand of



$5 \times 10^{-8}$  M (Fig. 2). The difference between our present findings and that of Park et al. [12] is obviously due to the expression system used and most likely to the G protein involved. This illustrates that it is important to choose the right expression system, when characterizing a GPCR. In this context, it is of interest to mention that CHO/G-16 cells are not always the optimal expression system for GPCR genes. For example, several years ago, we cloned the cDNA corresponding to the GPCR genes CG7285 and CG13702. When expressed in CHO/G-16 cells, we could not activate these receptors with any of the ligands contained in our insect neuropeptide library, including the *Drosophila* allatostatin-C (*Manduca sexta*-type allatostatin) peptide [16]. However, when expressed in frog oocytes, the CG7285/CG13702 receptors were activated by low concentrations of *Drosophila* allatostatin-C, showing that these receptors are allatostatin-C receptors [17]. Expression of the CG7285/CG13702 cDNAs, therefore, is more optimal in frog oocytes than in CHO/G-16 cells, the opposite situation as found for CG9918.

As mentioned in the Introduction, *Drosophila* produces three preprohormones that give rise to the seven related neuropeptides of Table 1. In addition to Drm-PK-1, three other *Drosophila* neuropeptides (Hug- $\gamma$ , Drm-PK-2, and Drm-ETH-1) activated the Drm-PK-1 receptor. However, six to eight times higher concentrations were needed to elicit the same effects as Drm-PK-1 (Table 1). Whether this cross-reactivity of the three neuropeptides with the Drm-PK-1 receptor is of physiological relevance is currently unclear. All three neuropeptides have the C-terminal sequence PRLamide in common with Drm-PK-1, which is probably the reason that they cross-reacted. It is interesting to note that it is not important whether the *Drosophila* peptides are pyrokinins (having the FXPRLamide sequence) or not. This is supported by our finding that leucopyrokinin (from the cockroach *L. maderae*) only showed a very low potency (Table 1). These results, then, show that the PRLamide C-terminal sequence is important for cross-reactivity, but that also the more N-terminal amino acid residues contribute. The most surprising result is that the Phe residue in the FXPRLamide consensus sequence apparently is not essential, or even can be a disadvantage (such as in the case of leucopyrokinin). These findings question the whole concept of the pyrokinin signature and it can be anticipated that when more PRLamide peptide derivatives will be tested on the Drm-PK-1 receptor, a better impression will be gained on the “true” pyrokinin consensus sequence.

All the above findings and conclusions are confirmed by a reinspection of our earlier work on the two Drm-PK-2 receptors (CG8795, CG8784) [11]. Although being a pyrokinin, Drm-PK-1 did hardly activate the two Drm-PK-2 receptors, whereas ETH-1 and Hug- $\gamma$ , although being non-pyrokinins, were about 100 times

more active [11]. Again, being a pyrokinin (having the FXPRLamide consensus sequence) can even be a disadvantage for the activation of a certain pyrokinin receptor. This conclusion, then, suggests that the “true” pyrokinin consensus sequence might, in fact, turn out to be two or more pyrokinin consensus sequences, each representing a certain pyrokinin subfamily that activates its own type of receptor.

## Acknowledgments

We thank Drs. S. Rees and J. Stables (Glaxo Wellcome, Stevenage, UK) for supplying cell line CHO/G-16, Sarah Dubédat for typing the manuscript, and the Danish Research Agency (Research Council for Nature and Universe), Lundbeck Foundation, Carlsberg Foundation, and Novo Nordisk Foundation (Fabrikant Vilhelm Pedersen og Hustrus Mindelegat) for financial support.

## References

- [1] M.D. Adams et al., Science 287 (2000) 2185–2195.
- [2] F. Hauser, M. Williamson, G. Cazzamali, C.J.P. Grimmelikhuijzen, Brief. Funct. Genom. Proteom. (2005), in press.
- [3] J. Vanden Broeck, Peptides 22 (2001) 241–254.
- [4] G.M. Holman, B.J. Cook, R.J. Nachman, Comp. Biochem. Physiol. C 85 (1986) 219–224.
- [5] R.J. Nachman, V.A. Roberts, H.J. Dyson, G.M. Holman, J.A. Tainer, Proc. Natl. Acad. Sci. USA 88 (1991) 4518–4522.
- [6] M. Altstein, Peptides 25 (2004) 1491–1501.
- [7] P. Verleyen, E. Clynen, J. Huybrechts, A. Van Lommel, L. Vanden Bosch, A. De Loof, J. Zdarek, L. Schoofs, Dev. Biol. 273 (2004) 38–47.
- [8] J. Zdarek, P. Verleyen, M. Mares, L. Doleckova, R.J. Nachman, J. Insect Physiol. 50 (2004) 233–239.
- [9] A. Iversen, G. Cazzamali, M. Williamson, F. Hauser, C.J.P. Grimmelikhuijzen, Biochem. Biophys. Res. Commun. 299 (2002) 924–931.
- [10] A. Iversen, G. Cazzamali, M. Williamson, F. Hauser, C.J.P. Grimmelikhuijzen, Biochem. Biophys. Res. Commun. 299 (2002) 628–633.
- [11] C. Rosenkilde, G. Cazzamali, M. Williamson, F. Hauser, L. Sondergaard, R. DeLotto, C.J.P. Grimmelikhuijzen, Biochem. Biophys. Res. Commun. 309 (2003) 485–494.
- [12] Y. Park, Y.J. Kim, M.E. Adams, Proc. Natl. Acad. Sci. USA 99 (2002) 11423–11428.
- [13] F. Staubli, T.J.D. Jørgensen, G. Cazzamali, M. Williamson, C. Lenz, L. Søndergaard, P. Roepstorff, C.J.P. Grimmelikhuijzen, Proc. Natl. Acad. Sci. USA 99 (2002) 3446–3451.
- [14] J. Stables, A. Green, F. Marshall, N. Fraser, E. Knight, M. Sautel, G. Miligan, M. Lee, S. Rees, Anal. Biochem. 252 (1997) 115–126.
- [15] U. Gether, Endocr. Rev. 21 (2000) 90–113.
- [16] M. Williamson, C. Lenz, Å.M.E. Winther, D.R. Nässel, C.J.P. Grimmelikhuijzen, Biochem. Biophys. Res. Commun. 282 (2001) 124–130.
- [17] H.J. Kreienkamp, H.J. Larusson, I. Witte, T. Roeder, N. Birgul, H.H. Honck, S. Harder, G. Ellinghausen, F. Buck, D. Richter, J. Biol. Chem. 277 (2002) 39937–39943.
- [18] F. Sadou, N. Amlaiki, J.L. Plassat, E. Borrelli, R. Hen, EMBO J. 9 (1990) 3611–3617.