



Neuropeptides and Neuropeptide Receptors in the *Drosophila melanogaster* Genome

Randall S. Hewes and Paul H. Taghert

Genome Res. 2001 11: 1126-1142

Access the most recent version at doi:[10.1101/gr.169901](https://doi.org/10.1101/gr.169901)

References This article cites 98 articles, 30 of which can be accessed free at:
<http://genome.cshlp.org/content/11/6/1126.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Neuropeptides and Neuropeptide Receptors in the *Drosophila melanogaster* Genome

Randall S. Hewes¹ and Paul H. Taghert²

Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110, USA

Recent genetic analyses in worms, flies, and mammals illustrate the importance of bioactive peptides in controlling numerous complex behaviors, such as feeding and circadian locomotion. To pursue a comprehensive genetic analysis of bioactive peptide signaling, we have scanned the recently completed *Drosophila* genome sequence for G protein-coupled receptors sensitive to bioactive peptides (peptide GPCRs). Here we describe 44 genes that represent the vast majority, and perhaps all, of the peptide GPCRs encoded in the fly genome. We also scanned for genes encoding potential ligands and describe 22 bioactive peptide precursors. At least 32 *Drosophila* peptide receptors appear to have evolved from common ancestors of 15 monophyletic vertebrate GPCR subgroups (e.g., the ancestral gastrin/cholecystokinin receptor). Six pairs of receptors are paralogs, representing recent gene duplications. Together, these findings shed light on the evolutionary history of peptide GPCRs, and they provide a template for physiological and genetic analyses of peptide signaling in *Drosophila*.

The recent publication of entire genomes for the worm, the fly, and human species has initiated the era of functional genomic analysis. The experiences to date have indicated that such analysis involves multiple stages, in which improvements are recorded as the databases are completed and analytic programs become more precise (Reese et al. 2000), and as more comparative information is made available (Sonnhammer et al. 1997). G protein-coupled receptors (GPCRs) provide sensitivity to a variety of environmental, developmental, and physiological signals. They display a uniform topology with seven transmembrane (TM) domains and represent one of the largest recognizable groups of proteins (Bockaert and Pin 1999). Here we have organized all genomic sequences that encode *Drosophila* GPCRs to identify and classify those devoted to peptide hormone and neuropeptide ligands (peptide GPCRs).

Given the availability of the human and mouse genomic sequences, what can we gain by a thorough analysis of *Drosophila* peptide GPCRs? We propose two reasons to motivate such efforts. First, our understanding of GPCR signaling mechanisms appears incomplete. Recent advances have indicated means by which GPCR signaling potential may be increased, including receptor oligomerization and association with a variety of accessory proteins (Bockaert and Pin 1999), and receptor translocation to the nucleus (Chen et al. 2000). There is great need, therefore, to address new hypotheses of GPCR signaling mechanisms in vivo. For this purpose, it will be very helpful to use the powerful

tools for genetic analysis that are afforded by model organisms such as *Drosophila*. The second reason we favor the pursuit of *Drosophila* GPCRs invokes the success of genetic analysis in another model genetic system, *Caenorhabditis elegans*. In the past few years, rapid progress has been made in the analysis of insulin signaling in the worm; *C. elegans* insulin regulates metabolism, development, and longevity by mechanisms that are similar to the endocrine regulation of metabolism and fertility by mammalian insulin (Kimura et al. 1997; Tissenbaum and Ruvkun 1998). In addition, the genetic analysis has extended our understanding of insulin signaling by revealing novel molecular features that may be variant in diabetic pedigrees (Ogg et al. 1997; Ogg and Ruvkun 1998). Although insulin binds to a different class of receptors, it is likely that the same rapid development of new information will accompany the genetic analysis of peptide GPCRs in *Drosophila*.

In a recent review, Brody and Cravchik (2000) began the process of categorizing *Drosophila* GPCRs by describing ~100 genes, including 21 receptors for classical neurotransmitters and neuromodulators (biogenic amines, related compounds, and purines) and 26–30 peptide receptor genes. We have extended that analysis by re-searching the original genomic sequences for peptide receptors (we found one additional GPCR) and by improving the annotations of 20 previously-predicted genes. We classified receptor genes according to phylogenetic trees constructed with the aid of the Pfam 7 TM databases (Bateman et al. 2000). In addition, we refined the *Drosophila* GPCR classifications by incorporating information we deduced by examining gene organizations. Through this analysis, we expanded the set of known and candidate peptide GPCRs from ~30 to ~45. Finally, to gain a sense of the

¹E-MAIL hewesr@thalamus.wustl.edu; FAX (314) 362-3446.

²E-MAIL taghertp@pcg.wustl.edu; FAX (314) 362-3446.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.169901>.

potential peptide ligands, we assembled a list of 22 *Drosophila* genes known or predicted to encode bioactive peptides that may activate these receptors. Together, these results shed light on the evolutionary history of neuropeptide signaling. They also are intended to aid in future efforts to analyze peptide receptor function in development, physiology, and behavior by using the power of *Drosophila* genetics.

RESULTS AND DISCUSSION

We searched the *Drosophila melanogaster* genome sequence with the goal of identifying all peptide GPCRs.

Initially, we scanned the gene annotations developed jointly by the Berkeley *Drosophila* Genome Project (BDGP) and Celera Genomics for all putative GPCRs (Adams et al. 2000; Brody and Cravchik 2000). Based on BLASTP scores obtained with each sequence, we excluded entries that are likely to represent nonpeptide GPCRs (neurexins, HE6- and methuselah-related proteins, rhodopsins, developmental genes, taste and odorant receptors, and receptors for biogenic amines and other small neurotransmitters). The remaining set of 44 known or putative peptide receptors and unclassified GPCRs was retained for further analysis (Table 1).

Table 1. Cloned and Candidate *Drosophila* Neuropeptide Receptors

Gene	Synonyms	Annotation corrected	Molecularly cloned	ESTs (#)
<i>AlstR</i>	<i>DAR-1</i> , <i>EG:121E7.2</i> , <i>CG2872</i>	no ^b	Birgöl et al. 1999; Lenz et al. 2000a	
<i>BG:BACR48G21.1</i>	—	n/a	no	
<i>CG1147</i>	—	yes	no	
<i>CG2114</i>	—	no	no	
<i>CG4187</i>	—	yes	no	
<i>CG4395</i>	—	yes	no	
<i>CG5042</i>	<i>CG5046</i>	yes	no	
<i>CG5811</i>	<i>NepYr</i> , <i>NYR</i> , <i>PR4</i>	no	Li et al. 1992b; St.-Onge et al. 2000	
<i>CG5911</i>	—	yes, splice variants	no	
<i>CG5936</i>	—	yes	no	
<i>CG6111</i>	—	no	no	
<i>CG6857</i>	—	yes	no	
<i>CG6881</i>	<i>CG6894</i>	yes	no	
<i>CG6986</i>	—	yes	no	
<i>CG7285</i>	—	yes	no	
<i>CG7395</i>	<i>CG18639</i>	yes	no	GH23382 (1)
<i>CG8422</i>	—	no	no	GH15162 (5)
<i>CG8784</i>	—	no	no	
<i>CG8795</i>	—	no	no	
<i>CG8985</i>	—	no, (RNA editing ^a)	no	HL01032 (1)
<i>CG9918</i>	—	no	no	
<i>CG10001</i>	<i>DAR-2</i>	no	Lenz et al. 2000b	
<i>CG10626</i>	—	no	no	
<i>CG10698</i>	—	no	no	
<i>CG10823</i>	—	yes	no	
<i>CG12370</i>	—	no	no	
<i>CG12610</i>	<i>CG15049</i> , <i>CG15050</i>	yes	no	
<i>CG13229</i>	—	no	no	
<i>CG13575</i>	—	no	no	
<i>CG13702</i>	—	yes	no	
<i>CG13758</i>	<i>EG:BACR25B3.3</i>	no	no	
<i>CG13803</i>	—	no	no	
<i>CG13995</i>	—	no	no	
<i>CG14003</i>	—	yes	no	
<i>CG14484</i>	<i>CG18192</i>	yes	no	
<i>CG14575</i>	—	yes	no	
<i>CG14593</i>	<i>CG14594</i>	yes	no	
<i>CG16726</i>	—	yes	no	
<i>CG17415</i>	—	yes	no	
<i>Fsh</i>	<i>DLGR-1</i> , <i>CG7665</i>	no	Hauser et al. 1997	
<i>GRHR</i>	<i>CG11325</i>	no	Hauser et al. 1998	GH20680 (3)
<i>rickets (rk)</i>	<i>DLGR-2</i> , <i>CG8930</i>	no ^b	Eriksen et al. 2000	
<i>Takr86C</i>	<i>NKD</i> , <i>CG6515</i>	no	Monnier et al. 1992	
<i>Takr99D</i>	<i>DTKR</i> , <i>CG7887</i>	no ^b	Li et al. 1991; 1992b	GH10154 (4)

EST, representative expressed sequence tag sequence (Rubin et al. 2000), with the number of ESTs given in parentheses.

^aEvidence of RNA editing from comparison of EST versus genomic DNA sequence (both from same genotype).

^bCG annotations incomplete, but complete sequences provided in independently published reports.

n/a, not applicable (not previously annotated).

To gauge the completeness of this set, we scanned the *Drosophila* genome for additional, nonannotated peptide GPCRs in three ways. First, we scanned a set of annotations of GPCR genes obtained through a GENSCAN search of the entire *Drosophila* genome (K. Scott and L. Vosshall, pers. comm.). This list contained one receptor sequence, *BG:BACR48G21.1* (BACR48G21.1), which had not been annotated previously. Second, we performed BLASTP searches using a “GPCR query set,” which included the previously cloned or annotated peptide, amine, and related/unclassified *Drosophila* GPCRs as well as ~200 sequences representing a diverse set of Family A GPCRs (unless indicated, we used the nomenclature for GPCR Families and Groups given by Kolakowski [1994] [see <http://www.gcrdb.uthscsa.edu/>]). All of the putative peptide receptors listed in Table 1 (including BACR48G21.1) were detected with several queries (for the vast majority, more than 30 times). However, this analysis did not reveal any additional candidate peptide GPCRs. Finally, we used the GPCR query set to perform TBLASTN searches of the Celera/BDGP whole genome shotgun sequence. As with the BLASTP survey, the TBLASTN search yielded scaffold sequences corresponding to all of the GPCRs on our list, but no additional candidate peptide receptors. The whole genome shotgun sequence currently represents ~98% of the *Drosophila* eukaryotic genome (Adams et al. 2000). Therefore, we conclude that the set of 45 cloned and candidate peptide GPCRs is essentially complete.

We focused on sequences encoding the seven TM domains. More than 50% of the BDGP/Celera annotations for GPCRs in this set (23 out of 43) were missing sequences representing one or more TM domains or, in the case of two receptors (CG4187 and CG5042), an N-terminal domain containing conserved, leucine-rich repeats (Table 1). In three cases, the correct gene sequences were published previously (Li et al. 1991; Ashburner et al. 1999; Birgül et al. 1999). We revised the remaining 20 incorrect or incomplete annotations using software-based gene prediction methods and manual inspection. For six GPCRs, there were two to three neighboring annotations that contained nonoverlapping GPCR sequence motifs. In each of these cases, we did not detect open reading frames encoding conserved TM domains in the intervening genomic sequence. Therefore, we merged these sequences to generate single, revised annotations. Additionally, for CG5911, we detected two adjacent sets of exons encoding alternative versions of TM4–7, including a conserved splice acceptor site in TM4. Thus, CG5911 appears to encode two distinct receptor isoforms (50% identical in TM5–7) through alternative splicing. Although final confirmation of these predictions will require direct sequencing of cDNAs, we conclude that our revised annotations are of sufficiently

high quality to perform phylogenetic analysis, based on the presence of well conserved motifs (Baldwin et al. 1997; Tams et al. 1998) throughout the TM domains of each of these receptors.

After assembling the list of 45 cloned and candidate peptide GPCRs, we classified these proteins based on BLASTP scores, on the locations of these receptors on representative phylogenetic trees for each GPCR Family (A and B), and on the degree to which these locations were supported by bootstrap analysis. We examined the genomic location of each peptide GPCR as well as all biogenic amine and small transmitter GPCRs to identify linked (possibly paralogous) genes. To detect conserved gene organizations, we noted the intron locations and phasing for each cloned and candidate peptide GPCR within the TM regions, as well as for a few related vertebrate GPCRs. The intron analysis of the vertebrate GPCRs was not comprehensive, in part because >93% of vertebrate GPCR genes lack introns within the coding sequence (Gentles and Karlin 1999). Finally, based on the results of the above tests, we were able to place most of the Family A receptors in one of seven alignments, each of which contained one or more related receptor subgroups.

We found strong evidence supporting the classification of 32 receptors as peptide GPCRs (Table 2). In addition, there were two receptors that are clear orthologs of the orphan receptor, LGR7, a member of a receptor clade containing several peptide GPCRs. For seven additional receptors, we found weaker evidence to indicate that they are peptide GPCRs. Finally, we regard four of the receptors as unclassifiable. Interestingly, we found at least six pairs of paralogs (variants generated by gene duplications or other processes), most of which appear to be related to common ancestors of vertebrate GPCR subgroups (rather than derived independently from vertebrate paralogs; e.g., Fig. 1A,D [see below]). Based on the presence of ESTs and/or cDNAs (Table 1), at least 25% of the genes in Table 1 are expressed. Pseudogenes are rare in *Drosophila* (Petrov and Hartl 2000), and based on strong sequence conservation in the TM domains, most of the remaining genes are likely to be expressed as well. Thus, we conclude that there are 39–41 peptide GPCRs in *Drosophila*, with an additional four GPCRs that may later be included in this category. The following sections describe this analysis in detail and provide a listing of potential cognate ligands.

Overview of the *Drosophila* Peptide GPCRs

Together, the set of known and candidate *Drosophila* peptide GPCRs contains representatives of at least 15 monophyletic vertebrate GPCR subgroups. Family A/Group III-B contains the largest number of *Drosophila* peptide GPCRs (at least 19; Table 2). These include 17 *Drosophila* representatives of six vertebrate

Table 2. Classification of Cloned and Candidate *Drosophila* Neuropeptide Receptors by BLAST and Phylogenetic Analysis

Gene	F/G	Best BLAST hit(s)				Phylogenetic analysis full tree neighbors	Bootstrap		
		Diagnostic	Pi	fly	Pi			Worm	P
CG6857	A/III-B	CCKR	35	CG6881	38	GASR, CCKR	>900		
CG6881	A/III-B	GASR	45	CG6857	71	GASR, CCKR	>900		
CG5811	A/III-B	GRL106	58	CG10626	48	GRL106	>700		
CG10626	A/III-B	LKR	77	CG5811	38	LKR, LSR	>990		
Takr86C	A/III-B	NK3R*	65	Takr99D	72	NKR	>500		
Takr99D	A/III-B	NK3R	64	Takr86C	67	NKR	>500		
CG10823	A/III-B	NFFR	45	CG5811	43	C50F7.1	38	OXR	<300
BG:BACR48G21.1	A/III-B	NK2R	20	Takr86C	21			NMBR/GRPR/BR	<100
CG1147	A/III-B	NY2R	31			C25G6.5	33	NPYR	<500
CG7395	A/III-B	NY2R	40			F41E7.3	32	NY2R	<500
CG12610	A/III-B	NPYR	14					PRPR	>500
CG13995	A/III	GASR	6					NMBR/GRPR/BR	<100
CG14593	A/III-B	BRS4	36	CG14484	52			NMBR/GRPR/BR	>700
CG14484	A/III-B	GRPR	40	CG14593	43			NMBR/GRPR/BR	>700
CG8784	A/III-B	NMUR	36	CG8795	117			NTR/GHSR	>700
CG8795	A/III-B	NMUR	39	CG8784	114			NTR/GHSR	>700
CG9918	A/III-B	NMUR	34	CG8784	61			NTR/GHSR	>700
CG14575	A/III-B	NMUR	42	CG8784	39	C48C5.1	39	NTR/GHSR	>700
CG14003	A/III?	GHSR	19	CG10001	20			NTR/GHSR	<100
CG5911A	A/III-B	TRFR	25					TRFR	<500
CG5911B	A/III-B	GHSR	28					ND	
CG2114	A/III?	TRFR	11			T14C1.1**	28	TRFR	<300
CG16726	A/III?	NK3R	14					TRFR	<300
CG6986	A/III?	TRFR	10	CG16726	11			TRFR	<300
CG13575	A/III	GHSR, LSR	8	CG10626	9			GPRJ	<300
CG8985	A/III?	AlstR	4	CG13803	130	F57B7.1**	29	TRFR	<300
CG13229	A/III?	NY5R	6	CG13803	46	F57B7.1**	31	TRFR	<300
CG13803	A/III?	AlstR	7	CG8985	130	F57B7.1**	25	TRFR	<300
CG5936	A/III?	NTR	2			B0563.6**	18	TRFR	<300
AlstR	A/V	GALR	41	CG10001	67	ZK455.3	43	GALR/GALS/GALT	<500
CG10001	A/V	GALR	28	AlstR	53	ZK455.3	31	GALR/GALS/GALT	<500
CG7285	A/V	SSR2	49	CG13702	78			SSR	<500
CG13702	A/V	SSR2	50	CG7285	75			SSR	<500
GRHR	A/V	GRHR	51					GRHR/VPR/OXYR	>700
CG10698	A/V	GRHR	36	GRHR	37			GRHR/VPR/OXYR	>700
CG6111	A/V	V1BR	38					GRHR/VPR/OXYR	>700
Fsh	A/V (c)	LSHR	66					FSHR/TSHR/LSHR/LGR4-6	>990
rk	A/V (c)	LSHR	68					FSHR/TSHR/LSHR/LGR4-6	>990
CG4187	A/V (c)	GRL101λ	57	CG5042	43			LGR7	>900
CG5042	A/V (c)	GRL101λ	34	CG4187	40			LGR7	>900
CG4395	B/I	CALR	22	CG17415	24			CALR/CGRR	>500
CG17415	B/I	CALR	44					CALR/CGRR	>500
CG13758	B/I-III	CALR	38			C13B9.4	50	B/I-III	<500
CG8422	B/I	DIHR	54	CG12370	65			DIHR	>900
CG12370	B/I	DIHR	71	CG8422	81			DIHR	>900

F/G, Family/Group classification. Diagnostic, Highest scoring receptor with known ligand/function. Fly, Highest scoring *Drosophila* GPCR (with *P* value near or exceeding the best diagnostic score). Worm, Highest scoring *Caenorhabditis elegans* GPCR (with *P* value near or exceeding the best diagnostic score). *P*, *P* value for respective diagnostic, fly or worm GPCR (10^{-x} , where *x* is the value displayed in the table).

**P* value for stable fly TKR is 74.

**Several other *C. elegans* GPCRs displayed *P* values exceeding the value obtained for the top "diagnostic" hit. CG10001 tends to be promiscuous on BLASTP with other, even very distantly related, fly receptors. λ, GRL101, as well as the other members of this class of GPCRs, is an orphan receptor (LGR7 is not in the public sequence databases).

ND, not done. (The remaining receptor abbreviations are described in the legends of Figs. 1–3.)

GPCR subgroups: the gastrin/cholecystokinin, neurokinin, neuropeptide FF and hypocretin/orexin, neuropeptide Y, bombesin/gastrin releasing peptide, and neurotensin receptors (and the neurotensin-related receptors for neuromedin U, growth hormone secretagogue, and thyrotropin releasing hormone). Family

A/Group V also contained a large number of *Drosophila* peptide GPCRs (11; Table 2), representing seven vertebrate GPCR subgroups: the galanin, somatostatin/opioid, gonadotropin releasing hormone, oxytocin/vasopressin, and glycoprotein hormone receptors, as well as two subgroups represented by vertebrate or-

phan receptors (LGR4–6 and LGR7). Finally, there were five *Drosophila* peptide GPCRs that belong to Family B. Four of these receptors belong to one of two vertebrate GPCR subgroups: the calcitonin and corticotropin releasing factor receptors. Thus, a large majority of the *Drosophila* and vertebrate neuropeptide signaling pathways appear to share common evolutionary origins. It remains to be seen whether the functions of these signals have been similarly conserved.

Family A/Group III-B: Gastrin/Cholecystokinin Receptors

Cholecystokinin (CCK) and gastrin are related neuroendocrine peptides that act through two closely related families of receptors (type A, CCKR, and type B, GASR). These receptors likely evolved from a common ancestor (Johnsen 1998). Two *Drosophila* GPCRs (CG6857, CG6881) displayed strong evidence of evolutionary kinship with this receptor subgroup (Table 2). On the subgroup-specific tree (Fig. 1A), CG6857 and CG6881 (as well as *Xenopus laevis* CCKR) were located on the base of the tree, before the branches leading to the CCKR and GASR receptors. Therefore, it appears that the fly receptors diverged from a common ancestor of the CCKR and GASR lineages. Consistent with this interpretation, CG6857 and CG6881 are closely linked genes (~30 kb apart), and they each display the strongest sequence similarity with each other (by BLASTP and on the phylogenetic trees; Table 2, Fig. 1A). Therefore, they likely arose through a gene duplication event rather than independently from the CCKR and/or GASR lineages. CG6857 and CG6881 both share an intron (same position and same phase) in TM3 (Table 3) with genes encoding both CCKR (accession #AF015959-AF015963) and human GASR (L10822). Likewise, all of these genes have an intron in a similar position within the highly variable cytoplasmic loop between TM5 and TM6. This

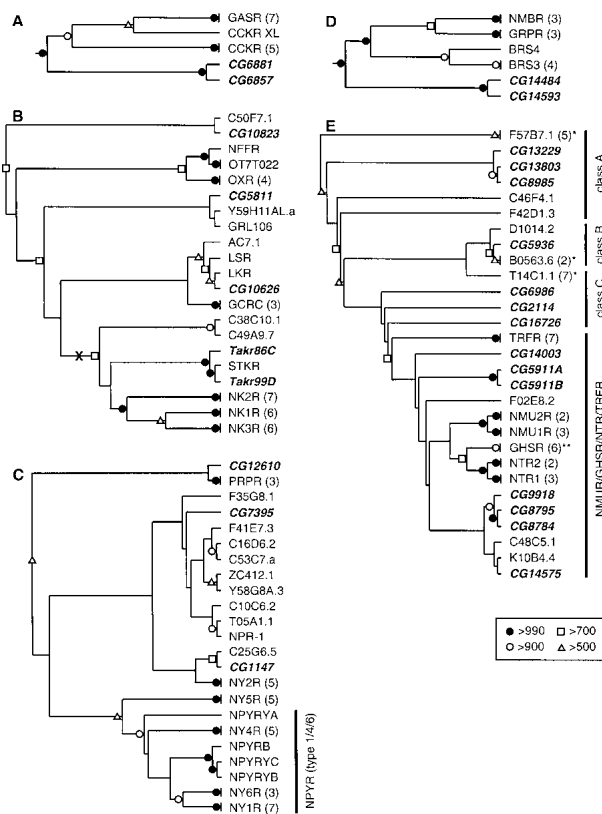


Figure 1 Neighbor-joining phylogenetic trees for the Family A, Group III-B receptors. For Family A and Group classifications, see Kolakowski (1994) (see <http://www.gcrdb.uthscsa.edu/>). (A) Rooted tree for the gastrin/cholecystokinin (CCK) receptors. (B) Unrooted tree for the neurokinin receptors (NKR) and related GPCRs. The midpoint of the tree is indicated with an "X." (C) Unrooted tree for the neuropeptide Y (NPY) receptors (NPYR) and prolactin releasing receptors (PRRs). (D) Rooted tree for the bombesin/gastrin releasing peptide receptors. (E) Unrooted tree for the neuromedin U receptors (NMURs), growth hormone secretagogue receptors (GHSRs), neurotensin receptors (NTRs), thyrotropin releasing hormone receptors (TRFRs), and a large family of related orphan receptors from *Caenorhabditis elegans*. The *C. elegans* orphan receptors included here belong to one of three clades (classes A–C). (*) Omitted

receptors are additional *C. elegans* orphan GPCRs; (**) omitted receptors are additional GHSRs and closely related orphan GPCRs. In A and D, a monophyletic set of 26 biogenic amine receptors (not shown) was used as the outgroup to determine the root of the trees (see Methods). In C and E, the location of the tree midpoint was ambiguous and is therefore not indicated. Portions of the trees representing groups of closely related receptors were omitted (the number of related receptors on each branch is indicated in parentheses). *Drosophila* GPCRs are listed in bold and italics. (BRS3) Bombesin receptor subtype 3; (BRS4) bombesin receptor subtype 4; (CCKR) CCK receptor type A; (CCKR XL) *Xenopus laevis* CCKR; (GASR) gastrin/CCK receptor type B; (GCRC) glucocorticoid-induced receptor; (GRL106) *Lymnaea stagnalis* cardioexcitatory receptor; (GRPR) gastrin releasing peptide (GRP) receptor; (LKR) *Boophilus microplus* (tick) leucokinin-like peptide receptor; (LSR) *L. stagnalis* lymnokinin receptor; (NFFR) neuropeptide FF/neuropeptide AF receptor; (NK1R–NK3R) NKR types 1–3; (NMU1R and NMU2R) neuromedin U receptor types 1 and 2; (NPR-1) product of the *C. elegans npr-1* gene; (NPYRYA–NPYRYC) orphan zebrafish NPYRs; (NPYRB) *Gadus morhua* (Atlantic cod) NPYR; (NTR1 and NTR2) neurotensin receptor types 1 and 2; (NY1R–NY6R) NPY receptor types 1–6; (OT7T022) putative mammalian RFRP receptor; (OXR) orexin/hypocretin receptor; (STKR) *Stomoxys calcitrans* (stable fly) tachykinin receptor. The remaining non-*Drosophila* sequences are orphan GPCRs from *C. elegans*. Symbols denote bootstrap support, out of 1000 replicates, that was >500: (filled circles) >990; (open circles) >900; (open squares) >700; (open triangles) >500.

conservation of introns further indicates that CG6857 and CG6881 are members of the CCKR/GASR receptor subgroup.

Family A/Group III-B: Neurokinin Receptors

The neurokinin (tachykinin) receptors (NKR) are a monophyletic group of GPCRs that are also closely related to the orexin/hypocretin receptors (OXRs), the neuropeptide FF/AF receptor (NFFR), and a class of orphan, glucocorticoid-induced receptors (GCRCs). We found six *Drosophila* members of this subgroup: CG5811 (Li et al. 1992b; St-Onge et al. 2000),

Table 3. Locations and Phasing of Introns among Genes Encoding *Drosophila* Family A Peptide GPCRs

	TM1		TM2		TM3		TM4		TM5		TM6		TM7
	after G	before D	after D	before P	after P	before P	after P	before P	after P	before P	after P	before P	
CG6857				24 (1)						82 (1)*		31 (2)	
CG6881				24 (1)			14 (1)			66 (2)			
CG5811					8 (2)		20 (2)				19 (0)	5 (0)	
CG10626			12 (0)		2 (2)		16 (1)				17 (0)		
Takr86C	10 (1)			7 (2)			11 (2)		13 (2)				
Takr99D		6 (1)		7 (2)					13 (2)			17 (1)	
CG10823		3 (0)	19 (1)	7 (2)									
BACR48G21.1				13 (0)									
CG1147		0 (1)		4 (0)									
CG7395													
CG12610				24 (0)		12 (0)	33 (0)	33 (1)			62 (2)*		
CG13995				24 (0)							57 (2)		
CG14593		9 (2)			6 (0)			14 (2)			8 (0)		8 (2)
CG14484		9 (2)			8 (1)		8 (0)				22 (0)	11 (2)	
CG8784			7 (1)		2 (2)		18 (0)				12 (1)		
CG8795			7 (1)		2 (2)		18 (0)				12 (1)		
CG9918											12 (1)		
CG14575			7 (1)	28 (2)			15 (1)				14 (2)		4 (2)
CG14003			19 (2)	14 (0)					6 (0)		79 (0)	9 (2)	
CG5911A/B	4 (0)		29 (1)			1 (2)							
CG2114													
CG16726								18 (0)	30 (2)	29 (1)	12 (0)		
CG6986	10 (2)			22 (0)		18 (1)		32 (0)		32 (0)			
CG13575													
CG8985													15 (1)
CG13229			2 (1)	2 (2)				15 (2)		29 (2)			
CG13803							1 (1)						15 (1)
CG5936		11 (1)		29 (2)	11 (2)								
AlstR	7 (2)	18 (2)				14 (2)		35 (0)		23 (2)	3 (0)		
CG10001							9 (0)				3 (0)		
CG7285										10 (0)	6 (0)		
CG13702										10 (0)			18 (0)
GRHR				7 (1)			2 (0)			28			
CG10698							2 (0)						20 (0)
CG6111		0 (1)			7 (1)				17 (1)				
Fsh	12 (2)		31 (1)						26 (1)				
rk									30 (2)	9 (0)			
CG4187			28 (1)		7 (2)					6 (2)	14 (2)	11 (1)?	
CG5042			23 (2)				19 (1)						

Family A peptide GPCRs are listed in the same order as shown in Table 2. The position of each intron is given with respect to the nearest amino acid landmark (position 0, underlined): TM1, GX_2GNXLV ; TM2, $NX_2NLAXADLL$; TM3, $SX_3LX_2ISXDRYX_2IX_2P$; TM4, $AX_7WX_2SX_3P$; TM5, $FX_2PLX_6YX_2I$; TM6, FX_2CWXP ; TM7, LX_3NSX_2NPXIY . Intron phase is specified as: |NNN phase 0; N|NN, phase 1; NN|N, phase 2 (where NNN is the codon and “|” is the position of the intron). Shaded items indicate shared introns. Dashed line, intron shared in two related GPCR subgroups.

*Only one out of multiple neighboring introns is indicated.

CG10626, TAKR86C (Monnier et al. 1992), TAKR99D (Li et al. 1991, 1992b), CG10823, and BACR48G21.1. On the subgroup-specific tree (Fig. 1B), TAKR86C and TAKR99D were located near the base of a branch leading to NK1R-NK3R, which indicates that the two fly proteins (and the two *C. elegans* orthologs) arose before the diversification the vertebrate NKRs. TAKR86C and TAKR99D are located together on the subgroup tree (with stable fly neurokinin receptor, STKR; Fig. 1B), and in BLASTP searches, each detects the other with the lowest *P* values (Table 2). Moreover, the *Takr86C* (Rosay et al. 1995) and *Takr99D* genes share two in-

trons in the same position and with the same phase (Table 3). Thus, TAKR86C and TAKR99D appear to be paralogs, and they are therefore likely to share similar ligands and functional properties.

Two additional *Drosophila* receptors, CG5811 and CG10626, are related to the true NKRs. However, based on BLASTP and phylogenetic analysis (Table 2, Fig. 1B), each of these receptors appears to be the ortholog of a NKR-related class of GPCRs that to date have been identified only in invertebrates. CG10626 is closely related to the tick NKR (LKR; Holmes et al. 2000) and the snail lymphokinin receptor (LSR) (Table 2), and these

three receptors are located on a single branch of the subgroup-specific tree (Fig. 1B). Likewise, CG5811 displays strong sequence similarity with GRL106, a snail NKR-like protein (Table 2). The branching pattern of this portion of the NKR subgroup tree is unstable (Fig. 1B). However, *CG5811* and *CG10626* have two introns that are in similar locations and display the same phasing (Table 3). Thus, these genes appear to be paralogs that diverged independently of the true NKRs. Consistent with this interpretation, the midpoint root of the NKR subgroup tree is located between the branch leading to the true NKRs and the branches of the tree leading to CG5811, CG10626, and the related GPCRs.

Finally, there are two additional GPCRs, CG10823 and BACR48G21.1, that display moderate to weak homology with the NKRs. By BLASTP, CG10823 displays strongest homology with the vertebrate neuropeptide FF/neuropeptide AF receptor (NFFR), the putative mammalian RF-amide-related peptide receptor (OT7T022; Hinuma et al. 2000), as well as CG5811 (Table 2). In addition, the *CG10823* gene has an intron that is located in the same position (and phase) as one of the two introns shared by *Takr86C* and *Takr99D* (Table 3). On the subgroup-specific tree, CG10823 is located near the base of a branch leading to the orexin/hypocretin receptors (OXRs), OT7T022 and NFFR (Fig. 1B). Therefore, CG10823 appears to have arisen from a common ancestor of these vertebrate relatives. Finally, BACR48G21.1 also appears to be a member of the NKR subgroup. However, this relationship was not well supported by the phylogenetic analysis (Table 2), and additional sequence data will be required to evaluate this finding.

Family A/Group III-B: Neuropeptide Y Receptors

The receptors for the neuropeptide Y (NPY) family of peptides (NPYRs) and the prolactin releasing peptide (PRPR) form a subgroup of related GPCRs (Hinuma et al. 1998; Hoyle 1999). Four *Drosophila* proteins, CG1147, CG7395, CG12610, and CG13995, appear to be members of this subgroup (Table 2). On the subgroup-specific tree, the position of the root was unclear (Fig. 1C). In addition, although the branching pattern for the portions of the tree containing the vertebrate NPYR receptors (except NY2R) was stable, the rest of the tree was not clearly resolved. In the BLASTP analysis and on the phylogenetic trees (Table 2; Fig. 1C), CG1147 showed the strongest sequence homology with a class of receptors that includes a *C. elegans* orphan GPCR (C25G6.5) and the vertebrate neuropeptide Y2 receptors (NY2Rs). CG7395, which also displays strong general sequence homology with the other members of this subgroup, appears to be most closely related to a diversified group of orphan NPYR-like *C. elegans* receptors. In contrast, CG12610 appears to be most closely related to PRPR. The fourth *Dro-*

sophila receptor in this group, CG13995, was located on the Group III portion of the full Family A tree, which consists almost exclusively of peptide GPCRs. However, CG13995 did not show strong evidence of homology with any specific class of peptide GPCRs (Table 2). Interestingly, the *CG13995* gene shares an intron in TM3 (same position and phase) with *CG12610*. Therefore, we propose that CG13995 is distantly related to the NPYR subgroup. Finally, it has been suggested that CG5811 is a NPYR-like receptor, despite its greater sequence similarity with the NKRs (see above), based on the activation of functionally expressed CG5811 by NPY and related peptides (at micromolar concentrations) and the lack of activation by vertebrate neurokinins (Li et al. 1992b). However, in competitive displacement experiments with CG5811 (St-Onge et al. 2000), PQGRF-amide-like peptides (e.g., NPF and *Lymnaea* cardioexcitatory peptide) displayed IC₅₀s in the subnanomolar range. Thus, CG5811 does not appear to be a member of the NPYR subgroup.

Family A/Group III-B: Bombesin/Gastrin Releasing Peptide Receptors

The bombesin-like neuropeptides, which include bombesin, gastrin releasing peptide (GRP), and neuromedin B (NMB), exert a wide variety of physiological actions in the CNS and the periphery through a class of related receptors (Sun et al. 2000). These receptors include the GRP-preferring receptor (GRPR), the neuromedin B-preferring receptor (NMBR), and an orphan class of receptors, characterized by bombesin receptor subtype 3 (BRS3). There are two *Drosophila* GPCRs, CG14484 and CG14593, that belong to this phylogenetic subgroup (Table 2). On the subgroup-specific tree, the three types of vertebrate bombesin/GRP receptors formed a clade, whereas CG14484 and CG14593 branch out from the base of the tree (Fig. 1D). Therefore, it appears that the fly receptors diverged from a common ancestor of the vertebrate bombesin/GRP receptor lineages. The organizations of the *CG14484* and *CG14593* genes are similar; each has one intron in the same position and phase, and there are two additional introns in similar positions (Table 3). Thus, CG14484 and CG14593 appear to be paralogs. Together, these results indicate that CG14484 and CG14593 are bombesin/GRP receptors; to our knowledge, this is the first clear molecular evidence for bombesin/GRP signaling in invertebrates.

Family A/Group III-B: Growth Hormone Secretagogue, Neurotensin, Neuromedin U, and Thyrotropin Releasing Hormone Receptors

The receptors for neurotensin (NTR), neuromedin U (NMUR), thyrotropin releasing hormone (TRFR), and growth hormone secretagogue (GHSR) form a large and diverse subgroup of GPCRs (Fujii et al. 2000). Among

these, NTR, GHSR, and NMUR display strong sequence similarity, whereas TRFR is more distantly related. At least seven *Drosophila* GPCRs appear to be members of this subgroup: CG8784, CG8795, CG9918, CG14575, CG5911A, CG5911B, and CG14003 (Table 2). An additional seven GPCRs (CG2114, CG5936, CG6986, CG8985, CG13229, CG13803, and CG16726) are all most closely related to a large set of related orphan receptors that had been identified previously only in *C. elegans* (C. Bargmann, pers. comm.). These orphan GPCRs fall into at least three classes, and there are one

to three *Drosophila* GPCRs in each class (Fig. 1E). The three receptors in class A (CG8985, CG13229, and CG13803) display strong sequence homology. In addition, CG8985 and CG13803 are linked genes (~30 kb apart), and they share an intron (Table 3). Thus, the *Drosophila* class A receptors appear to be paralogs. All three classes display weak sequence similarity with TRFR, NTR, and GHSR, indicating that this entire family of orphan receptors may be derived from an ancestor of these vertebrate receptors and therefore may encode peptide GPCRs. However, confirmation of such a relationship will require functional analysis of one or more members of these orphan GPCR classes.

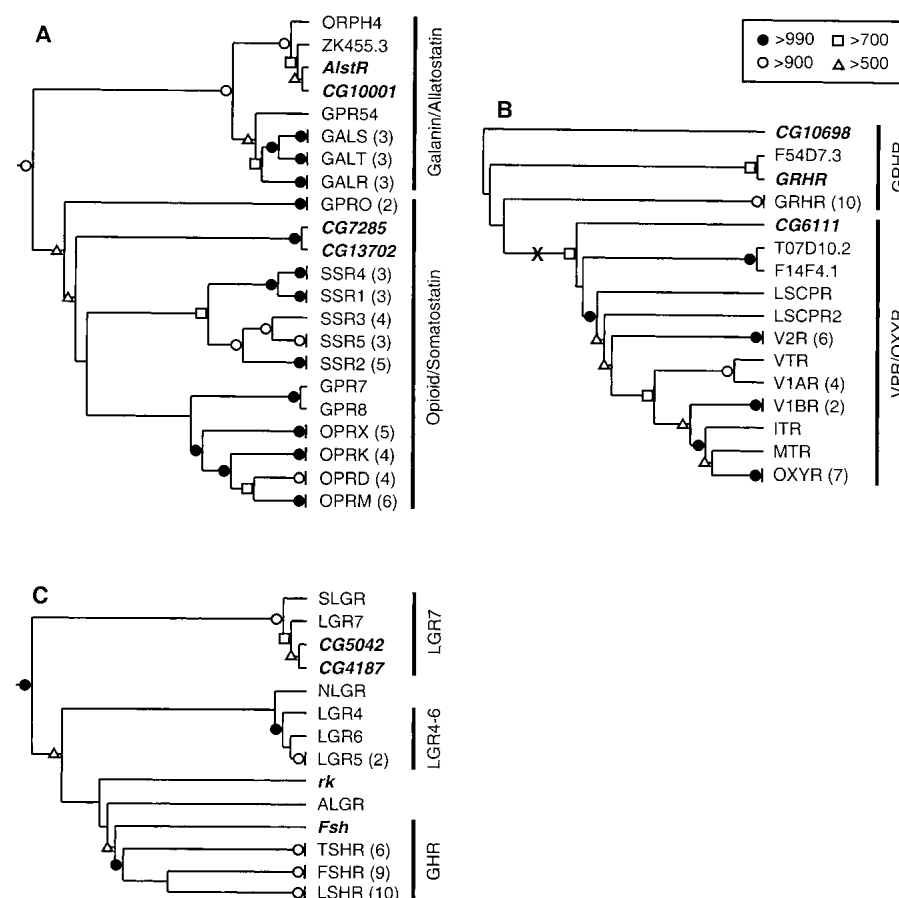


Figure 2 Neighbor-joining phylogenetic trees for the Family A, Group V receptors. (A) Rooted tree for the opioid, somatostatin, galanin, and allatostatin receptors. (B) Unrooted tree for the gonadotropin releasing hormone (GnRH), vasopressin, and oxytocin receptors. The likely mid-point of the tree is indicated with an "X." (C) Rooted tree for the glycoprotein hormone receptors and related leucine-rich repeat containing receptors (LGRs). Bootstrap scores, omitted branches, and *Drosophila* GPCRs are indicated as in Fig. 1. (ALGR) *Anthopleura elegantissima* (sea anemone) LGR; (FSHR) follicle-stimulating hormone receptor; (GALR) galanin receptor type 1; (GALS) galanin receptor type 2; (GALT) galanin receptor type 3; (GPR24 and GPR54) mammalian orphan GPCRs; (GRHR) GnRH receptor; (ITR) isotocin receptor; (LGR4–7) LGR types 4–7; (LSCPR and LSCPR2) *Lymanaea stagnalis* conopressin receptor types 1 and 2; (LSHR) lutropin-choriogonadotropic hormone receptor; (MTR) mesotocin receptor; (NLGR) *C. elegans* LGR; (ORPH4) *Lymanaea stagnalis* orphan GPCR; (OPRD) delta-type opioid receptor; (OPRK) kappa-type opioid receptor; (OPRM) mu-type opioid receptor; (OPRX) nociceptin/orphanin FQ receptor; (OXYR) oxytocin receptor; (SLGR) *L. stagnalis* GRL101; (SSR1–SSR5) somatostatin receptor types 1–5; (TSHR) thyrotropin receptor; (V1AR and V1BR) vasopressin V1A and V1B receptors; (V2R) vasopressin V2 receptor; (VTR) vasotocin receptor. The remaining non-*Drosophila* sequences are orphan GPCRs from *C. elegans*.

relationship will require functional analysis of one or more members of these orphan GPCR classes.

CG8784 and CG8795 are two of the seven *Drosophila* GPCRs displaying the strongest sequence similarity with this vertebrate subgroup, and they appear to be paralogs. They display strong sequence similarity with each other (Table 2; Fig. 1E). Moreover, the CG8784 and CG8795 genes are closely linked (~10 kb apart) and share four introns with identical positions and phasing (Table 3). Similarly, CG9918 and CG14575 each share one intron with CG8784/CG8795 (Table 3), indicating that all four of these receptors are closely related. Their closest vertebrate homologs are NMUR, GHSR, and NTR, based on BLASTP analysis and on their positions in the phylogenetic trees (Table 2; Fig. 1E). Consistent with this finding, the shared intron located in the TM6 domain of CG8784 and CG8795 is also found in the same position and with the same phasing in the pufferfish GHSR gene (AF082211). However, the branching pattern for the subgroup-specific tree was unstable, and the evolutionary relationships among these receptors are unclear. Three additional receptors, CG5911A and CG5911B (generated by putative alternative splicing of the CG5911 gene) and CG14003, also displayed moderate to

weak sequence homology with this subgroup and appear to be most closely related to vertebrate TRFR.

Family A/Group V: Galanin/Allatostatin and Opioid/Somatostatin Receptors

There were four *Drosophila* receptors, *AlstR* (Birgöl et al. 1999; Lenz et al. 2000a), CG7285, CG10001 (Lenz et al. 2000b), and CG13702, that displayed strong sequence similarity with galanin, somatostatin, and opioid receptors (Table 2). Because these three classes of vertebrate receptors display extensive sequence similarity, we grouped them together to construct a subgroup-specific tree (Fig. 2A). The root of this tree is located between the branch leading to the galanin receptors and the branch leading to the somatostatin and opioid receptors. CG7285 and CG13702 were located on the branch containing all of the somatostatin and opioid receptors and related orphan GPCRs. The opioid receptors form a clade, and two groups of somatostatin receptors also form clades (SSR1/4 in one and SSR2/3/5 in the other). The remaining branches on this side of the tree are unstable. Together, these results indicate that CG7285 and CG13702 are orthologous to the vertebrate somatostatin and opioid receptors, although it is not clear whether they diverged from a common ancestor or from a point deeper within the tree. CG7285 and CG13702 appear to be paralogs; they display strong sequence homology (Table 2; Fig. 2A), and they are encoded by linked genes (~90 kb apart) that share an intron with the same location and phasing (Table 3).

The allatostatin receptor, *AlstR* (Birgöl et al. 1999), and CG10001 were located on the portion of the tree containing all of the galanin receptors (Fig. 2A), indicating that *AlstR* and CG10001 are *Drosophila* orthologs of the mammalian galanin receptors. This finding is in agreement with an earlier phylogenetic analysis of *AlstR* (Birgöl et al. 1999). The *AlstR* and *CG10001* genes share an intron at the same location and with the same phasing (Table 3; Lenz et al. 2000b). Thus, *AlstR* and *CG10001* appear to be paralogs and are likely to share many functional properties. Interestingly, immunocytochemical studies, using anti-porcine galanin and anti-porcine galanin message-associated peptide, as well as receptor autoradiography studies using ¹²⁵I-porcine galanin, showed the presence of galanin-like peptides in several locations in the adult CNS of blowflies, including the fan-shaped body of the central complex and a ring of cells in the medulla (Lundquist et al. 1991, 1993; Johard et al. 1992). Similar patterns of staining in the fan-shaped body and medulla have been obtained in *Drosophila* with a specific monoclonal anti-allatostatin antiserum (Yoon and Stay 1995). These comparative data provide additional support for the conclusion that *AlstR* and CG10001 are closely related to the ver-

tebrate galanin receptors (cf., Birgöl et al. 1999; Lenz et al. 2000b).

Family A/Group V: Gonadotropin Releasing Hormone, Vasopressin, and Oxytocin Receptors

The receptors for gonadotropin releasing hormone (GRHR) and the receptors for vasopressin (VPR) and oxytocin (OXYR) belong to two closely related clades of GPCRs (Hoyle 1999). In *Drosophila*, there are three GPCRs that belong to this subgroup; CG6111, CG10698, and Dm-GRHR (Table 2; Hauser et al. 1998). The branching pattern near the base of the subgroup-specific tree was unstable (Fig. 2B), and the evolutionary history of this subgroup is unclear. However, when the tree is midpoint rooted, Dm-GRHR and CG10698 branch from the side of the tree leading to the vertebrate GRHRs, and CG6111 branches from the side of the tree leading to VPR, OXYR, and related GPCRs. These results are in agreement with the results of BLASTP analysis. Moreover, the *Dm-GRHR* gene shares an intron near TM4 (identical location and phasing) with the rat *GRHR* gene (U92471)(Hauser et al. 1998); *CG10698* also shares this intron. Thus, *Drosophila* appears to have two GRHR-like receptors and one VPR/OXYR-like receptor.

Family A/Group V (Type 1c): Glycoprotein Hormone Receptors

Four glycoprotein hormones have been identified in mammals: thyroid-stimulating hormone (TSH) and the gonadotropins, follicle-stimulating hormone (FSH), choriogonadotropin (CG), and luteinizing hormone (LH). These four hormones bind to a subgroup of receptors (the LGRs) that all bear a characteristic, large, N-terminal "ectodomain" that participates in the binding of the large glycoprotein ligands (Hsu et al. 2000) (type 1c receptors; Bockaert and Pin 1999). Four *Drosophila* receptors, CG4187, CG5042, and the proteins encoded by the *Fsh* (Hauser et al. 1997) and *rk* (Ashburner et al. 1999; Eriksen et al. 2000) genes, display sequence similarity with the LGRs, including the N-terminal ectodomain (Table 2). On the subgroup-specific tree (Fig. 2C), there were three distinct clades (cf., Hsu et al. 2000). The first includes LGR7, a *Lymnaea* ortholog (SLGR), CG4187, and CG5042. The second includes LGR4–LGR6, and the third includes the glycoprotein hormone receptors (LSHR, FSHR, and TSHR). *Fsh* is located at the base of a branch leading to the glycoprotein hormone receptors, indicating that this gene may have evolved from a common ancestor of LSHR, FSHR, and TSHR. Three additional receptors, *C. elegans* LGR (NLGR), sea anemone LGR (ALGR), and *rk*, were grouped only weakly with the glycoprotein hormone receptors; the branching pattern of this portion of the tree was unstable. Therefore, these could

not be assigned to any one class of LGRs by basis of the phylogenetic analysis alone.

Within the ectodomain, all of the LGRs contain a variable number of leucine-rich repeats and a functionally important hinge region located between the leucine-rich repeats and the seven-TM core. At the borders of the hinge region, there are two sequences that are diagnostic of the three different subclasses of LGRs (Table 4; Hsu et al. 2000). These groupings are also supported by BLASTP analysis of the ectodomains (data not shown). These sequences support the placement of *Fsh* in the subfamily of glycoprotein hormone receptors.

Placement of CG4187 and CG5042 in the LGR7 clade is supported by BLASTP analysis of the ectodomains (data not shown) and the presence of subgroup-specific hinge sequences (Table 4). Unlike the other two subgroups of LGRs, the ectodomains of LGR7 and snail LGR have low density lipoprotein (LDL) receptor-like cysteine-rich motifs at the N terminus (Tensen et al. 1994; Hsu et al. 2000). CG4187 and CG5042 also each contain at least one LDL motif (Table 4). The function of the LDL motif is unclear, but it indicates a possible role for lipoprotein-like molecules in neuronal G protein-mediated signal transduction (Tensen et al. 1994). Alternatively, given the presence of leucine-rich repeats, these receptors may bind to glycoproteins.

Although phylogenetic analysis of the LGRs did not place *rk* in any of the three subgroups of LGRs, analysis of the ectodomain indicates that this receptor

is orthologous to LGR4–6. This is based on the presence of hinge sequences most similar to LGR4–6 and on BLASTP analysis (data not shown). The other members of this family are orphan receptors. However, the presence of the leucine-rich repeats indicates that these proteins also bind to glycoproteins.

Family B/Group I: Calcitonin and Diuretic Hormone Receptors

In addition to the 40 proteins in Family A (the rhodopsin-like receptors), there are 5 *Drosophila* peptide GPCRs in Family B (the secretin-like receptors). Based on BLASTP analysis and their positions on the phylogenetic tree (Fig. 3), at least four of these receptors (CG4395, CG8422, CG12370, and CG17415) belong to Group I. This group contains the receptors for calcitonin (CALR), calcitonin gene related peptide (CGRR), corticotropin releasing factor (CRFR and CRF2), and diuretic hormone (DIHR). The position of the fifth *Drosophila* peptide GPCR in this family (CG13758) is unclear, and it may be a member of Group I, II, or III. CG8422 and CG12370 appear to be paralogs, and they are orthologous to the DIHRs. These receptors belong to a clade containing CRFR and CRF2, which indicates that the ancestor to the insect DIHRs evolved from a common ancestor of the vertebrate corticotropin releasing factor receptors (Fig. 3). In contrast, CG4395 and CG17415 are most closely related to CALR and CGRR, although the bootstrap scores more deeply located within this branch of the tree were not strong enough to determine whether CALR and CGRR diverged before or after the related *Drosophila* receptors. We did not find evidence for well defined GPCR-associated proteins (e.g., RAMPs [Bockaert and Pin 1999] and RCPs [Evans et al. 2000]).

Drosophila Genes Encoding Neuropeptides and Peptide Hormones

We wished to compare the number of peptide GPCRs with the number of neuropeptides present (or suspected to exist) in *Drosophila*. Based on the literature and on some genomic analysis, we have assembled a list of 22 *Drosophila* neuropeptide genes (Table 5). These genes are either known or predicted to encode bioactive neuropeptides and peptide hormones. Eight of these, which encode neuropeptides described for *Drosophila* or other arthropods, were described previously only in gene annotations generated by Celera/BDGP and in a parallel survey, which was just published recently (Vanden Broeck 2001). An additional peptide listed by Vanden Broeck (2001) (“IFa”) was not included, because the precursor did not match our criteria for putative neuropeptide genes. Because neuropeptide-encoding precursors do not display multiple, uniform characteristics found in GPCRs, we are certain to have missed many peptide genes and thus consider

Table 4. Conserved LGR Hinge Sequences

LGR4–6 consensus	YAYQCC	GXFKPCEX
<i>rk</i>	YAYHCC	GPFLPCAD
LGR4	YAYQCC	GAFKPCEY
LGR5	YAYQCC	GPFPKCEH
FEX	SAYQCC	GPFPKCEH
FSHR/LSHR/TSHR consensus	YPSHCC	DXFNPCED
<i>Fsh</i>	HSFHCC	NDLNPCED
NLGR	YPHHCC	DALNPCEN
FSHR_human	YPSHCC	DAFNPCED
LSHR_human	YPSHCC	DAFNPCED
TSHR_human	YPSHCC	DEFNPCED
ALGR	NGFLCC	DAFHPCED
LGR7 consensus	XZXZCX	DGZSSXXX
CG4187	NVRVCD	DGISSKLH
CG5042	RFFYCS	DGVSSFQD
LGR7	KFYQCG	DGISSLEN
SLGR	SYRFCC	DEFSSCED
LDL receptor motif		
CG4187 ^a		NCDGSVDCDDASDEVNC
CG5042		CVPRRQMCDSRNDCADSSDENPVEC

The consensus hinge sequences for LGR4–6 and FSHR/LSHR/TSHR were described by Hsu et al. (2000). The consensus hinge sequences for LGR7 are based on a ClustalX alignment of these four proteins.

^aThis sequence is truncated N-terminally, presumably due to missing sequence at the N-terminal end of the annotation.

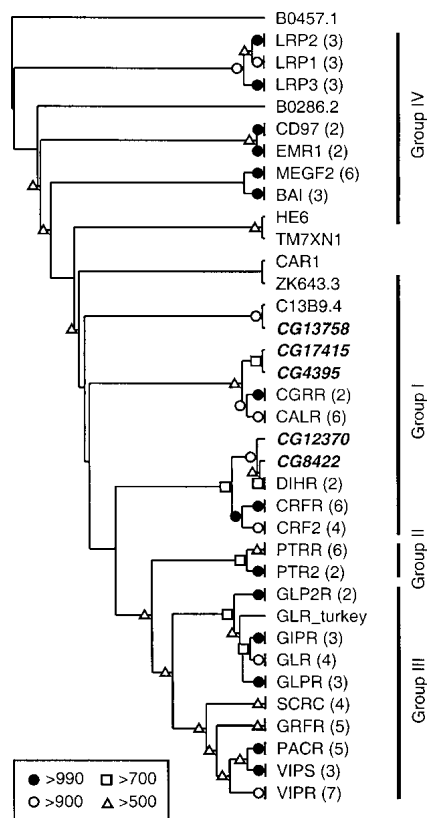


Figure 3 Unrooted neighbor-joining tree for the Family B receptors. The location of the tree midpoint is ambiguous and is therefore not indicated. Bootstrap scores, omitted branches, and *Drosophila* GPCRs are indicated as in Fig. 1. The four groups of Family B receptors are indicated with vertical bars. (BAI) brain-specific angiogenesis inhibitors 1–3; (CALR) calcitonin receptor; (CAR1) cyclic AMP receptor 1; (CD97) leucocyte antigen CD97; (CGRR) calcitonin gene-related peptide type 1 receptor; (CRF2) corticotropin releasing factor (CRF) receptor 2; (CRFR) CRF receptor 1; (DIHR) diuretic hormone receptor; (EMR1) cell surface glycoprotein EMR1; (GIPR) gastric inhibitory polypeptide receptor; (GLP2R) glucagon-like peptide 2 receptor; (GLPR) glucagon-like peptide 1 receptor; (GLR) glucagon receptor; (GRFR) growth hormone releasing hormone receptor; (HE6) G protein-coupled receptor HE6; (LRP1–3) calcium-independent alpha-latrotoxin receptors (latrophilins) 1–3; (MEGF2) seven-pass transmembrane proteins CELSR1–2 and MEGF2; (PACR) pituitary adenylate cyclase activating polypeptide (PACAP) type I receptor; (PTR2) parathyroid hormone receptor; (PTRR) parathyroid hormone/parathyroid hormone-related peptide receptor; (SCRC) secretin receptor; (TM7XM1) human EGF-TM7 like protein; (VIPR) vasoactive intestinal polypeptide (VIP) receptor 1; (VIPS) VIP receptor 2. The remaining non-*Drosophila* sequences are orphan GPCRs from *Caenorhaloditis elegans*.

this list incomplete. However, assuming a 1 : 1 ratio of neuropeptide and peptide hormone genes to peptide GPCRs, these 22 genes appear to encode the ligands for at least 50% of the *Drosophila* peptide GPCRs that we have described. This may be an underestimate, given the fact that many of these neuropeptide genes encode multiple peptides. In addition to these 22 neuropeptide genes, we list several insect peptides and peptide

hormones known in other insects and for which *Drosophila* homologs have been inferred by observation or simply by conjecture. Although the structures of these genes are not yet available, they are included here to permit consideration of all plausible ligands for the identified peptide GPCRs.

Ligands for Family A Peptide GPCRs

There are multiple genes encoding potential ligands for the *Drosophila* NKR-like receptors. CG14734 produces neurokinin-like peptides (Siviter et al. 2000) that likely bind to TAKR86C and TAKR99D, as shown by functional expression of these receptors and specific binding to mammalian (Li et al. 1991) and insect (Monnier et al. 1992) neurokinins and related peptides. Based on the pharmacological observations of the *Lymnaea* lymnokinin receptor, LSR (Cox et al. 1997), we speculate that the *Drosophila* ortholog, CG10626, is a receptor for the leucokinin-like peptides encoded by CG13480 (Terhzaz et al. 1999).

In addition to the neurokinin-like peptides, there are also several genes that are known to encode (or potentially encode) peptides terminating in the sequence RF-amide. These include the putative peptide products of a novel gene, CG13968. Along with the products of the *dFMRFa* (Nambu et al. 1988; Schneider and Taghert 1988) and *DMS* (CG6440) genes, these peptides may interact with multiple receptors for RF-amide peptides. CG5811 has been shown to bind with high affinity to molluscan -PQGRF-amide peptides and therefore is likely to represent the first of several *Drosophila* RF-amide peptide receptors (St-Onge et al. 2000). A second potential RF-amide peptide receptor, CG10823, is orthologous to NFFR and OT7T022, both of which bind ligands bearing the C-terminal consensus sequence PXRf-amide (Elshourbagy et al. 2000; Hinuma et al. 2000). The other *Drosophila* neuropeptides ending in RF-amide are found within the *disk* gene and display structural similarity to the vertebrate cholecystokinins (Nichols et al. 1988). We speculate that these peptides may interact with either or both of the paralogous CCKR/GASR-related receptors (CG6857 and CG6881).

AlstR binds a native *Drosophila* allatostatin peptide (AST-1) with high affinity (Birgöl et al. 1999). This peptide, along with multiple other related peptides, is encoded by CG13633 (Lenz et al. 2000c). Because AlstR and CG10001 are paralogs, the latter receptor is also likely to interact with one or more of the products of CG13633.

Ligands for Family B Peptide GPCRs

We speculate that the corticotropin releasing factor (CRF)-related peptides encoded by CG8348 and CG13094 (similar to *Locusta* diuretic hormone; Coast

Table 5. *Drosophila* Neuropeptides and Peptide Hormones

Peptide name	Gene	Location	EST	Precursor	Molecularly cloned	Related sequence(s)/evidence for gene
amnesiac	CG11937	19A1	LP07893	170	Feany and Quinn 1995	
diuretic hormone	CG13094	29D1		116		Furuya et al. 2000
M-ASH	CG14919	32D2		122		Kramer et al. 1991
LRLRFamide	CG13968	38B4		300		Feng et al. 1999
dFMRFamide	CG2346	46C2		347	Nambu et al. 1988; Schneider and Taghert 1988	
ion transport peptide	CG13586	60D5		430		Lee et al. 1995; Meredith et al. 1996
ETH/P-ETH	CG18105	60D16		203	Park et al. 1999	
AKH	CG1171	64A10		79	Noyes et al. 1995	
sex peptide	CG17673	70A4		55	Kubli 1992; Ottiger et al. 2000	
leucokinin	CG13480	70E3		153	Terhzaz et al. 1999	
allatostatin (B-type)	CG6456	74B1	GH13904	211	Williamson et al. 2001	
drosulfakinin	CG18090	82A1		128	Nichols et al. 1988	
diuretic hormone	CG8348	85E2	GH27214	>183		Coast 1996
neurokinin	CG14734	87A9		289	Siviter et al. 2000	
corazonin	CG3302	88B7	LP11439	72	Veenstra 1994	
NP-PP	CG10342	89D	GH04563	102	Brown et al. 1999	
EH	CG5400	90B1		97	Horodyski et al. 1993	
CCAP	CG4910	94C4		151		Cheung et al. 1992; Lehman et al. 1993; Ewer and Truman 1996
dromyosuppressin	CG6440	95F1	GH10451	100		Nichols 1992
allatostatin	CG13633	96A23		151	Lenz et al. 2000c	
PDF	CG6496	97B2		102	Park and Hall 1998	
CAP2B/pyrokinin	CG15520	99D1	GH28004	148		Huesmann et al. 1995; Davies et al. 1998
allatotropin	ND					Kataoka et al. 1989; Žitňan et al. 1993
anterior retraction factor, pupariation tanning factor	ND					Sivasubramanian et al. 1974
baratin	ND					Nässel et al. 2000
bursicon	ND					Fraenkel et al. 1966; Kostron et al. 1999
colliculostatin, NEB, TMOF	ND					De Loof et al. 1995
ecdysiotropin	ND					Koolman et al. 1995
GBP	ND					Hayakawa et al. 1995
neuroparsin, parsin	ND					Girardie et al. 1998
PBAN	ND					Kawano et al. 1992; Sato et al. 1993; Masler et al. 1994; Zdarek et al. 1997
proctolin	ND					Anderson et al. 1988
PTTH (large)	ND					Kim et al. 1997
somatostatin-like	ND					Ui-Tei et al. 1995
steroidogenic factor	ND					Brown et al. 1998
vasopressin-like	ND					Baines et al. 1995

(Location) cytological location determined directly or inferred from the location of the sequenced genomic clones. (EST) expressed sequence tag clone. (Precursor) known or deduced length of the prepropeptide. (Molecularly cloned), gene identification verified by cDNA (excluding ESTs) and/or in situ hybridization. (ND) not determined. (Related sequence(s)/evidence for gene) existence of neuropeptide gene inferred from peptide sequences from *Drosophila* or other arthropods, physiological assays, and/or immunocytochemical data. An additional predicted neuropeptide gene (*Ifa*, Vanden Broeck 2001) lacks structural features.

1996; Furuya et al. 2000) interact with the CRFR-related CG8422 and/or CG12370, both of which are orthologs (Fig. 3) of the *Acheta domesticus* diuretic hormone receptor (Reagan 1994). Additionally, Zhong and Pena (1995) found evidence for a PACAP-like pep-

tide in flies. Feany and Quinn (1995) and Moore et al. (1998) provided genetic evidence to implicate the *amnesiac* gene (potentially encoding peptides of the PACAP family) in various *Drosophila* behaviors. We speculate that the peptides in this group may interact

with one or more of the remaining Family B receptors (CG4395, CG17415, and CG13758).

Peptide Genes Still Awaiting Identification

There are several insect neuropeptides and peptide hormones that have not as yet been cloned in *Drosophila*. These include three large protein hormones—PTTH, bursicon, and the anterior retraction factor (ARF)—that are known to exist in *Drosophila* but currently lack molecular definition. At least two of these proteins, PTTH and bursicon, are glycoprotein hormones (Fraenkel et al. 1966; Kim et al. 1997), whereas the structure of ARF remains undefined (Sivasubramanian et al. 1974). As noted above, the structure of the receptor encoded by the *fsH* gene indicates that it binds to a glycoprotein hormone ligand. All of the mammalian glycoprotein hormones share a similar structure, consisting of common α - and specific β -subunits (Hsu et al. 2000). However, to date, no similar proteins have been identified in flies. We speculate that PTTH, bursicon, and ARF are good candidate ligands for members of the LGR class of receptors.

Relatives of several peptides identified in other insects may also be present in *Drosophila*. These include PBAN and diapause hormone (DH), which are found in diverse insects. Both are peptide hormones of moderate size that are cosynthesized along with shorter pyrokinin peptides (Kawano et al. 1992; Sato et al. 1993; Masler et al. 1994; Xu et al. 1995). It is notable that the *Drosophila* CG15520 precursor includes a single FXPRL-amide (pyrokinin-like) peptide but lacks any sequences similar to PBAN or DH. Because NMUR is activated by peptides displaying a LXXPRX-amide consensus (Fujii et al. 2000), we speculate that this pyrokinin-like peptide may interact with CG14484 and/or CG14593, which are orthologs of NMUR. Likewise, the *Drosophila* ecdysis triggering hormones (ETHs), which have a PRX-amide C-terminal sequence (Park et al. 1999), may signal through these receptors.

With the completion of the *D. melanogaster* genome sequence, we are now able to take a comprehensive picture of the genes encoding peptide GPCRs in this species, and a complete catalog of the cognate ligands should soon follow. This is an important first step toward detailed physiological and genetic analyses of neuropeptide signaling in *Drosophila*.

METHODS

Peptide GPCR Sequence Acquisition

To identify all predicted *Drosophila* GPCRs, we first scanned the gene annotations developed jointly by the Berkeley *Drosophila* Genome Project (BDGP) and Celera Genomics for all proteins predicted to contain domains matching seven-TM motifs (Adams et al. 2000; Brody and Cravchik 2000). We rejected sequences identified recently as odorant receptors by a committee representing scientists working in the field (*Dro-*

sophila Odorant Receptor Nomenclature Committee 2000). Each remaining cloned and candidate receptor gene was used as a BLASTP search query of the database of predicted *Drosophila* proteins using the BDGP server (<http://www.fruitfly.org/>) and/or of the “non-redundant” database of all proteins using the NCBI server (<http://www.ncbi.nlm.nih.gov/>). Sequences were not considered further if the resulting top-scoring proteins yielded *P* values for nonpeptide receptors (and associated orphan receptors) that were at least 10-fold greater than the top *P* value for a putative peptide receptor. Three sequences (CG18314, CG12796, CG13579) generated a smaller range of *P* values following BLASTP searches on the BDGP server. Nevertheless, analysis of these proteins using the NCBI server yielded hits that were exclusively amine/small neurotransmitter receptors (or orphan receptors). These proteins therefore are likely to encode nonpeptide receptors, and they also were excluded.

We first scanned a set of GPCR sequence annotations obtained through a GENSCAN search of the complete *Drosophila* genome sequence (see Vossell et al. 1999) and identified based on sequence similarity to GPCRs in the NCBI nonredundant protein database (K. Scott and L. Vossell, pers. comm.). For the BLASTP and TBLASTN analyses, we assembled a “GPCR query set,” which included the previously annotated peptide, amine, and related/unclassified *Drosophila* GPCRs as well as ~200 sequences representing a diverse set of Family A (rhodopsin receptor-like family) GPCRs from the Pfam database (7TM-1; <http://pfam.wustl.edu/>). These sequences were used as queries for BLASTP and TBLASTN searches on the BDGP server, using the predicted proteins and the Celera/BDGP whole genome shotgun sequence datasets, respectively. To expedite the latter search, we assumed that TBLASTN hits to genomic sequences that were already on our list were due to the detection of previously annotated GPCR genes.

GPCR Alignments

We used the hidden Markov model–based protein alignments contained in Version 5.5 of PFAM (Sept., 2000; Bateman et al. 2000) as a template for the manual alignment of the *Drosophila* cloned and candidate peptide receptors. The alignments were viewed using CLUSTALX (Version 1.8; Thompson et al. 1997), and, in some cases, this program was used to help resolve the alignment of variable regions (e.g., between TM domains 4 and 5). We used these alignments to build phylogenetic trees and also to detect missing or incorrect sequences in the gene annotations.

The N-terminal and C-terminal non-TM sequences in GPCRs tend to be poorly conserved, making accurate alignment difficult, and the seven-TM core region is sufficient for the subclassification of these proteins (Strader et al. 1994). Therefore, for Family A receptors, we deleted sequences N-terminal to the conserved GNXXLV motif (single-letter amino acid code) in TM1 and C-terminal to the conserved NPXIY motif in TM7. For Family B receptors (secretin receptor family), we deleted sequences flanking the X₁₀GX₃S motif in TM1 and the QGX₂V X₄CX₅X motif in TM7.

Correction of Annotations

To locate missing TM domains among the putative peptide receptor annotations, we scanned for potential coding exons in flanking genomic sequence using the GENSCAN server at MIT (<http://genes.mit.edu/GENSCAN.html>), and the FGENES (gene prediction) and FEX (exon prediction) programs on the

Baylor College of Medicine (BCM Search Launcher) server (<http://www.hgsc.bcm.tmc.edu/>). We also scanned for potential mRNA splice sites using the SPL program on the BCM Search Launcher server and by manual inspection of potential open reading frames displayed using MacVector (Genetics Computer Group, Madison, WI). The DNA sequences for all of the predicted donor and acceptor splice sites were NN|GT and AG|NN, respectively. Finally, we examined neighboring gene annotations to identify duplicate annotations of single GPCR genes. The annotations were judged to be complete when each of the TM domains displayed features that were clearly recognizable among closely related receptors. Except for the LGR subgroup of receptors (see Results and Discussion), which all share a large and subgroup-specific N-terminal domain, we did not evaluate the quality of the annotations for the N-terminal and C-terminal non-TM regions.

Tree Building

We classified the cloned and candidate peptide GPCRs based on five criteria. First, we noted the highest scoring BLASTP hits obtained on the NCBI server (Table 2). Second, we constructed alignments of Family A and Family B receptors, including all of the *Drosophila* peptide GPCRs identified above, for the purpose of generating full phylogenetic trees for each family. For Family A, we included mostly complete (TM1–TM7) sequences representing each of the five receptor groups, as well as sequences representing each of the various subgroups of receptors (e.g., the three types of galanin receptors) and representative orphan receptors contained within the full list of Pfam 7TM-1 (Family A) GPCRs. For Family B, we included all of the Group I–III receptors and a representative set of Family B, Group IV receptors within the full list of Pfam 7TM-2 GPCRs. After manual editing of the alignments, we constructed neighbor-joining phylogenetic trees for each family using ClustalX, using the correction for multiple substitutions provided by the software, followed by bootstrap analysis (1000 replicates).

For the subsequent subgroup-specific trees, we attempted to include all complete TM1–TM7 sequences belonging to each subgroup (as well as some partial sequences). These were identified by scanning the full Pfam 7TM-1 alignment and the GPCRDB listing of available GPCR sequences (<http://www.gpcr.org/7tm/>), and by performing BLASTP searches with the cloned and candidate *Drosophila* peptide GPCRs as well as other representatives of each subgroup. After manual editing of the alignments, the construction of neighbor-joining trees and the bootstrap analysis was performed as above. A set of 26 indoleamine (biogenic amine) receptors, which form a monophyletic group (Kolakowski 1994), was used as an outgroup for the purpose of rooting the subgroup-specific trees. When the position of the root was unclear, the outgroup was omitted. All alignments, revised annotations, and unabridged versions of the trees are located at <http://thalamus.wustl.edu/flyGPCR/peptideGPCR.html>. In addition, the revised annotations have been submitted to FlyBase (<http://flybase.bio.indiana.edu/>).

NOTE ADDED IN PROOF

We have identified two additional ESTs for peptide GPCRs: AT008361 (CG1147) and AT0019640 (CG13229). Also, based on discussions with Jan Veenstra (Universite Bordeaux) we now add two additional peptide genes to our list: the *SIFamide*

gene (currently listed as part of *CG4681*; Ifa, Vanden Broeck 2001) and the *hugin* gene (*CG6371*).

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant NS21749 and the Human Frontier Science Program Organization (P.H.T.). We thank Sean Eddy for helpful discussions, Kirstin Scott and Leslie Vossball for sharing *Drosophila* GPCR sequence data, Lin Yang and Dori Sztipanovits for technical assistance, and Aguan Wei for comments on the manuscript. We also thank Cori Bargmann and Kemal Payza for sharing unpublished results.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Anderson, M.S., Halpern, M.E., and Keshishian, H. 1988. Identification of the neuropeptide transmitter proctolin in *Drosophila* larvae: Characterization of muscle fiber-specific neuromuscular endings. *J. Neurosci.* **8**: 242–255.
- Ashburner, M., Misra, S., Roote, J., Lewis, S.E., Blazej, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N., et al. 1999. An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*. The *Adh* region. *Genetics* **153**: 179–219.
- Baines, R.A., Thompson, K.S., Rayne, R.C., and Bacon, J.P. 1995. Analysis of the peptide content of the locust vasopressin-like immunoreactive (VPLI) neurons. *Peptides* **16**: 799–807.
- Baldwin, J.M., Schertler, G.F., and Unger, V.M. 1997. An α -carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* **272**: 144–164.
- Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L.L. 2000. The Pfam protein families database. *Nucleic Acids Res.* **28**: 263–266.
- Birgöl, N., Weise, C., Kreienkamp, H.-J., and Richter, D. 1999. Reverse physiology in *Drosophila*: Identification of a novel allatostatin-like neuropeptide and its cognate receptor structurally related to the mammalian somatostatin/galanin/opioid receptor family. *EMBO J.* **18**: 5892–5900.
- Bockaert, J. and Pin, J.P. 1999. Molecular tinkering of G protein-coupled receptors: An evolutionary success. *EMBO J.* **18**: 1723–1729.
- Brody, T. and Cravchik, A. 2000. *Drosophila melanogaster* G protein-coupled receptors. *J. Cell Biol.* **150**: F83–F88.
- Brown, M.R., Graf, R., Swiderek, K.M., Fendley, D., Stracker, T.H., Champagne, D.E., and Lea, A.O. 1998. Identification of a steroidogenic neurohormone in female mosquitoes. *J. Biol. Chem.* **273**: 3967–3971.
- Brown, M.R., Crim, J.W., Arata, R.C., Cai, H.N., Chun, C., and Shen, P. 1999. Identification of a *Drosophila* brain-gut peptide related to the neuropeptide Y family. *Peptides* **20**: 1035–1042.
- Chen, R., Mukhin, Y.V., Garnovskaya, M.N., Thielen, T.E., Iijima, Y., Huang, C., Raymond, J.R., Ullian, M.E., and Paul, R.V. 2000. A functional angiotensin II receptor-GFP fusion protein: Evidence for agonist-dependent nuclear translocation. *Am. J. Physiol. Renal. Physiol.* **279**: F440–F448.
- Cheung, C.C., Loi, P.K., Sylwester, A.W., Lee, T.D., and Tublitz, N.J. 1992. Primary structure of a cardioactive neuropeptide from the tobacco hawkmoth, *Manduca sexta*. *FEBS Lett.* **313**: 165–168.
- Coast, G.M. 1996. Neuropeptides implicated in the control of diuresis in insects. *Peptides* **17**: 327–336.
- Cox, K.J., Tensen, C.P., Van der Schors, R.C., Li, K.W., van

- Heerikhuizen, H., Vreugdenhil, E., Geraerts, W.P., and Burke, J.F. 1997. Cloning, characterization, and expression of a G-protein-coupled receptor from *Lymnaea stagnalis* and identification of a leucokinin-like peptide, PSFHSWSamide, as its endogenous ligand. *J. Neurosci.* **17**: 1197–1205.
- Davies, S.A., Stewart, E.J., Huesmann, G.R., Skaer, N.J., Maddrell, S.H., Tublitz, N.J., and Dow, J.A. 1998. Neuropeptide stimulation of the nitric oxide signaling pathway in *Drosophila melanogaster* Malpighian tubules. *Am. J. Physiol.* **273**: R823–R827.
- De Loof, A., Bylemans, D., Schoofs, L., Janssen, I., Spittaels, K., Vanden Broeck, J., Huybrechts, R., Borovsky, D., Hua, Y.J., Koolman, J., et al. 1995. Folliculostatin, gonadotropins and a model for control of growth in the grey fleshfly, *Neobellieria (sarcophaga) bullata*. *Insect Biochem. Mol. Biol.* **25**: 661–667.
- Drosophila* Odorant Receptor Nomenclature Committee. 2000. A unified nomenclature system for the *Drosophila* odorant receptors. *Cell* **102**: 145–146.
- Elshourbagy, N.A., Ames, R.S., Fitzgerald, L.R., Foley, J.J., Chambers, J.K., Szekeres, P.G., Evans, N.A., Schmidt, D.B., Buckley, P.T., Dytko, G.M., et al. 2000. Receptor for the pain modulatory neuropeptides FF and AF is an orphan G protein-coupled receptor. *J. Biol. Chem.* **275**: 25965–25971.
- Eriksen, K.K., Hauser, F., Schiott, M., Pedersen, K.M., Sondergaard, L., and Grimmelikhuijzen, C.J. 2000. Molecular cloning, genomic organization, developmental regulation, and a knock-out mutant of a novel leu-rich repeats-containing G protein-coupled receptor (DLGR-2) from *Drosophila melanogaster*. *Genome Res.* **10**: 924–938.
- Evans, B.N., Rosenblatt, M.I., Mnayer, L.O., Oliver, K.R., and Dickerson, I.M. 2000. CGRP-RCP, a novel protein required for signal transduction at calcitonin gene-related peptide and adrenomedullin receptors. *J. Biol. Chem.* **275**: 31438–31443.
- Ewer, J. and Truman, J.W. 1996. Increases in cyclic 3', 5'-guanosine monophosphate (cGMP) occur at ecdysis in an evolutionarily conserved crustacean cardioactive peptide-immunoreactive insect neuronal network. *J. Comp. Neurol.* **370**: 330–341.
- Feany, M.B. and Quinn, W.G. 1995. A neuropeptide gene defined by the *Drosophila* memory mutant *amnesiac*. *Science* **268**: 869–873.
- Feng, G., Reale, V., Kennedy, K., Chatwin, H.M., Evans, P.D., and Hall, L.M. 1999. Cloning and functional characterization of a novel neuropeptide F-like receptor from *Drosophila melanogaster*. *Soc. Neurosci. Abst.* **29**: 183.
- Fraenkel, G., Hsiao, C., and Seligman, M. 1966. Properties of bursicon: an insect protein hormone that controls cuticular tanning. *Science* **151**: 91–93.
- Fujii, R., Hosoya, M., Fukusumi, S., Kawamata, Y., Habata, Y., Hinuma, S., Onda, H., Nishimura, O., and Fujino, M. 2000. Identification of neuromedin U as the cognate ligand of the orphan G protein-coupled receptor FM-3. *J. Biol. Chem.* **275**: 21068–21074.
- Furuya, K., Milchak, R.J., Schegg, K.M., Zhang, J., Tobe, S.S., Coast, G.M., and Schooley, D.A. 2000. Cockroach diuretic hormones: Characterization of a calcitonin-like peptide in insects. *Proc. Natl. Acad. Sci.* **97**: 6469–6474.
- Gentles, A.J. and Karlin, S. 1999. Why are human G-protein-coupled receptors predominantly intronless? *Trends Genet.* **15**: 47–49.
- Girardie, J., Huet, J.C., Atay-Kadiri, Z., Ettaouil, S., Delbecq, J.P., Fournier, B., Pernollet, J.C., and Girardie, A. 1998. Isolation, sequence determination, physical and physiological characterization of the neuroparsins and ovary maturing parsins of *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* **28**: 641–650.
- Hauser, F., Nothacker, H.-P., and Grimmelikhuijzen, C.J.P. 1997. Molecular cloning, genomic organization, and developmental regulation of a novel receptor from *Drosophila melanogaster* structurally related to members of the thyroid-stimulating hormone, follicle-stimulating hormone, luteinizing hormone/choriogonadotropin receptor family from mammals. *J. Biol. Chem.* **272**: 1002–1010.
- Hauser, F., Sondergaard, L., and Grimmelikhuijzen, C.J. 1998. Molecular cloning, genomic organization and developmental regulation of a novel receptor from *Drosophila melanogaster* structurally related to gonadotropin-releasing hormone receptors for vertebrates. *Biochem. Biophys. Res. Commun.* **249**: 822–828.
- Hayakawa, Y., Ohnishi, A., Yamanaka, A., Izumi, S., and Tomino, S. 1995. Molecular cloning and characterization of cDNA for insect biogenic peptide, growth-blocking peptide. *FEBS Lett.* **376**: 185–189.
- Hinuma, S., Habata, Y., Fujii, R., Kawamata, Y., Hosoya, M., Fukusumi, S., Kitada, C., Masuo, Y., Asano, T., Matsumoto, H., et al. 1998. A prolactin-releasing peptide in the brain. *Nature* **393**: 272–276.
- Hinuma, S., Shintani, Y., Fukusumi, S., Iijima, N., Matsumoto, Y., Hosoya, M., Fujii, R., Watanabe, T., Kikuchi, K., Terao, Y., et al. 2000. New neuropeptides containing carboxy-terminal RFamide and their receptor in mammals. *Nat. Cell. Biol.* **2**: 703–708.
- Holmes, S.P., He, H., Chen, A.C., Ivie, G.W., and Pietrantonio, P.V. 2000. Cloning and transcriptional expression of a leucokinin-like peptide receptor from the southern cattle tick, *Boophilus microplus* (Acari: Ixodidae). *Insect Mol. Biol.* **9**: 457–465.
- Horodyski, F.M., Ewer, J., Riddiford, L.M., and Truman, J.W. 1993. Isolation, characterization and expression of the eclosion hormone gene of *Drosophila melanogaster*. *Eur. J. Biochem.* **215**: 221–228.
- Hoyle, C.H.V. 1999. Neuropeptide families and their receptors: Evolutionary perspectives. *Brain Res.* **848**: 1–25.
- Hsu, S.Y., Kudo, M., Chen, T., Nakabayashi, K., Bhalla, A., van der Spek, P.J., van Duin, M., and Hsueh, A.J.W. 2000. The three subfamilies of leucine-rich repeat-containing G protein-coupled receptors (LGR): Identification of LGR6 and LGR7 and the signaling mechanism for LGR7. *Mol. Endocrinol.* **14**: 1257–1271.
- Huesmann, G.R., Cheung, C.C., Loi, P.K., Lee, T.D., Swiderek, K.M., and Tublitz, N.J. 1995. Amino acid sequence of CAP2b, an insect cardioacceleratory peptide from the tobacco hawkmoth *Manduca sexta*. *FEBS Lett.* **371**: 311–314.
- Johard, H.A., Lundquist, C.T., Rokaesus, A., and Nässel, D.R. 1992. Autoradiographic localization of ¹²⁵I-galanin binding sites in the blowfly brain. *Regul. Pept.* **42**: 123–134.
- Johnsen, A.H. 1998. Phylogeny of the cholecystokinin/gastrin family. *Front. Neuroendocrinol.* **19**: 73–99.
- Kataoka, H., Toschi, A., Li, J.P., Carney, R.L., Schooley, D.A., and Kramer, S.J. 1989. Identification of an allatropin from adult *Manduca sexta*. *Science* **243**: 1481–1483.
- Kawano, T., Kataoka, H., Nagasawa, H., Isogai, A., and Suzuki, A. 1992. cDNA cloning and sequence determination of the pheromone biosynthesis activating neuropeptide of the silkworm, *Bombyx mori*. *Biochem. Biophys. Res. Commun.* **189**: 221–226.
- Kim, A.J., Cha, G.-H., Kim, K., Gilbert, L.I., and Lee, C.C. 1997. Purification and characterization of the prothoracicotropic hormone of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **94**: 1130–1135.
- Kimura, K.D., Tissenbaum, H.A., Liu, Y., and Ruvkun, G. 1997. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **15**: 942–946.
- Kolakowski, L.F. 1994. GCRDb: A G-protein-coupled receptor database. *Receptors Channels* **2**: 1–7.
- Koolman, J., Hua, Y.-J., Byelans, D., and De Loof, A. 1995. A potential prothoracostatic hormone (PTSH) from flies. In *Molecular mechanisms of insect metamorphosis and diapause* (eds. A. Suzuki et al.), pp. 45–54. Industrial Pub., Tokyo.
- Kostron, B., Market, D., Kellermann, J., Carter, C.E., and Honegger, H.W. 1999. Antisera against *Periplaneta americana* Cu,Zn-superoxide dismutase (SOD): Separation of the neurohormone bursicon from SOD, and immunodetection of SOD in the central nervous system. *Insect Biochem. Mol. Biol.* **29**: 861–871.
- Kramer, S.J., Toschi, A., Miller, C.A., Kataoka, H., Quistad, G.B., Li, J.P., Carney, R.L., and Schooley, D.A. 1991. Identification of an allatostatin from the tobacco hornworm *Manduca sexta*. *Proc. Natl. Acad. Sci.* **88**: 9458–9462.
- Kubli, E. 1992. The sex-peptide. *Bioessays* **14**: 779–784.
- Lee, K.J., Elton, T.S., Bej, A.K., Watts, S.A., and Watson, R.D. 1995.

- Molecular cloning of a cDNA encoding putative molt-inhibiting hormone from the blue crab, *Callinectes sapidus*. *Biochem. Biophys. Res. Commun.* **209**: 1126–1131.
- Lehman, H.K., Murguic, C.M., Miller, T.A., Lee, T.D., and Hildebrand, J.G. 1993. Crustacean cardioactive peptide in the sphinx moth, *Manduca sexta*. *Peptides* **14**: 735–741.
- Lenz, C., Sondergaard, L., and Grimmlikhuijzen, C.J. 2000a. Molecular cloning and genomic organization of a novel receptor from *Drosophila melanogaster* structurally related to mammalian galanin receptors. *Biochem. Biophys. Res. Commun.* **269**: 91–96.
- Lenz, C., Williamson, M., and Grimmlikhuijzen, C.J. 2000b. Molecular cloning and genomic organization of a second probable allatostatin receptor from *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **273**: 571–577.
- . 2000c. Molecular cloning and genomic organization of an allatostatin preprohormone from *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **273**: 1126–1131.
- Li, X.J., Wolfgang, W.J., Wu, Y.N., North, R.A., and Forte, M. 1991. Cloning, heterologous expression and developmental regulation of a *Drosophila* receptor for tachykinin-like peptides. *EMBO J.* **10**: 3221–3229.
- Li, X.J., Wu, Y.N., North, R.A., and Forte, M. 1992b. Cloning, functional expression, and developmental regulation of a neuropeptide Y receptor from *Drosophila melanogaster*. *J. Biol. Chem.* **267**: 9–12.
- Lundquist, C.T., Rokaeus, A., and Nässel, D.R. 1991. Galanin immunoreactivity in the blowfly nervous system: Localization and chromatographic analysis. *J. Comp. Neurol.* **312**: 77–96.
- Lundquist, C.T., Johard, H.A., Rokaeus, A., and Nässel, D.R. 1993. Galanin immunoreactivity and ¹²⁵I-galanin binding sites in the blowfly brain. *Acta. Biol. Hung.* **44**: 51–54.
- Masler, E.P., Raina, A.K., Wagner, R.M., and Kochansky, J.P. 1994. Isolation and identification of a pheromonotropic neuropeptide from the brain-suboesophageal ganglion complex of *Lymantria dispar*: A new member of the PBAN family. *Insect Biochem. Mol. Biol.* **24**: 829–836.
- Meredith, J., Ring, M., Macins, A., Marschall, J., Cheng, N.N., Theilmann, D., Brock, H.W., and Phillips, J.E. 1996. Locust ion transport peptide (ITP): Primary structure, cDNA and expression in a baculovirus system. *J. Exp. Biol.* **199**: 1053–1061.
- Monnier, D., Colas, J.F., Rosay, P., Hen, R., Borrelli, E., and Maroteaux, L. 1992. NKD, a developmentally regulated tachykinin receptor in *Drosophila*. *J. Biol. Chem.* **267**: 1298–1302.
- Moore, M.S., DeZazzo, J., Luk, A.Y., Tully, T., Singh, C.M., and Heberlein, U. 1998. Ethanol intoxication in *Drosophila*: Genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**: 997–1007.
- Nambu, J.R., Murphy-Erdosh, C., Andrews, P.C., Feistner, G.J., and Scheller, R.H. 1988. Isolation and characterization of a *Drosophila* neuropeptide gene. *Neuron* **1**: 55–61.
- Nässel, D.R., Persson, M.G., and Muren, J.E. 2000. Baratin, a nonamidated neurostimulating neuropeptide, isolated from cockroach brain: Distribution and actions in the cockroach and locust nervous systems. *J. Comp. Neurol.* **422**: 267–286.
- Nichols, R. 1992. Isolation and structural characterization of *Drosophila* TDVDHVFLRFamide and FMRFamide-containing neural peptides. *J. Mol. Neurosci.* **3**: 213–218.
- Nichols, R., Schneuwly, S.A., Dixon, J.E. 1988. Identification and characterization of a *Drosophila* homologue to the vertebrate neuropeptide cholecystokinin. *J. Biol. Chem.* **263**: 12167–12170.
- Noyes, B.E., Katz, F.N., and Schaffer, M.H. 1995. Identification and expression of the *Drosophila* adipokinetic hormone gene. *Mol. Cell. Endocrinol.* **109**: 133–141.
- Ogg, S. and Ruvkun, G. 1998. The *C. elegans* PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. *Mol. Cell* **2**: 887–893.
- Ogg, S., Paradis, S., Gottlieb, S., Patterson, G.I., Lee, L., Tissenbaum, H.A., and Ruvkun, G. 1997. The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**: 994–999.
- Ottiger, M., Soller, M., Stocker, R.F., and Kubli, E. 2000. Binding sites of *Drosophila melanogaster* sex peptide pheromones. *J. Neurobiol.* **44**: 57–71.
- Park, J.H. and Hall, J.C. 1998. Isolation and chronobiological analysis of a neuropeptide pigment-dispersing factor gene in *Drosophila melanogaster*. *J. Biol. Rhythms* **13**: 219–228.
- Park, Y., Zitňan, D., Gill, S.S., and Adams, M.E. 1999. Molecular cloning and biological activity of ecdysis-triggering hormones in *Drosophila melanogaster*. *FEBS Lett.* **463**: 133–138.
- Petrov, D.A. and Hartl, D.L. 2000. Pseudogene evolution and natural selection for a compact genome. *J. Hered.* **91**: 221–227.
- Reagan, J.D. 1994. Expression cloning of an insect diuretic hormone receptor. A member of the calcitonin/secretin receptor family. *J. Biol. Chem.* **269**: 9–12.
- Reese, M.G., Hartzell, G., Harris, N.L., Ohler, U., Abril, J.F., and Lewis, S.E. 2000. Genome annotation assessment in *Drosophila melanogaster*. *Genome Res.* **10**: 483–501.
- Rosay, P., Colas, J.F., Maroteaux, L. 1995. Dual organisation of the *Drosophila* neuropeptide receptor NKD gene promoter. *Mech. Dev.* **51**: 329–339.
- Rubin, G.M., Yandell, M.D., Wortman, J.R., Miklos, G.G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., et al. 2000. Comparative genomics of the eukaryotes. *Science* **287**: 2204–2215.
- Sato, Y., Oguchi, M., Menjo, N., Imai, K., Saito, H., Ikeda, M., Isobe, M., and Yamashita, O. 1993. Precursor polyprotein for multiple neuropeptides secreted from the suboesophageal ganglion of the silkworm *Bombyx mori*: Characterization of the cDNA encoding the diapause hormone precursor and identification of additional peptides. *Proc. Natl. Acad. Sci.* **90**: 3251–3255.
- Schneider, L.E., and Taghert, P.H. 1988. Isolation and characterization of a *Drosophila* gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH₂ (FMRFamide). *Proc. Natl. Acad. Sci. USA* **85**: 1993–1997.
- Sivasubramanian, P., Friedman, S., and Fraenkel, G. 1974. Nature and role of proteinaceous hormonal factors acting during puparium formation in flies. *Biol. Bull.* **147**: 163–185.
- Siviter, R.J., Coast, G.M., Winther, A.M., Nachman, R.J., Taylor, C.A., Shirras, A.D., Coates, D., Isaac, R.E., and Nässel, D.R. 2000. Expression and functional characterization of a *Drosophila* neuropeptide precursor with homology to mammalian preprotachykinin A. *J. Biol. Chem.* **275**: 23273–23280.
- Sonnhammer, E.L., Eddy, S.R., and Durbin, R. 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins* **28**: 405–420.
- St-Onge, S., Fortin, J.-P., Labarre, M., Steyaert, A., Schmidt, R., Ahmad, S., Walker, P., and Payza, K. 2000. In vitro pharmacology of NPFF and FMRFamide-related peptides at the PR4 receptor of *Drosophila melanogaster*. *Soc. Neurosci. Abstr.* **26**: 140.9.
- Strader, C.D., Fong, T.M., Tota, M.R., and Underwood, D. 1994. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* **63**: 101–132.
- Sun, B., Schally, A.V., and Halmos, G. 2000. The presence of receptors for bombesin/GRP and mRNA for three receptor subtypes in human ovarian epithelial cancers. *Regul. Pept.* **90**: 77–84.
- Tams, J.W., Knudsen, S.M., and Fahrenkrug, J. 1998. Proposed arrangement of the seven transmembrane helices in the secretin receptor family. *Receptors Channels* **5**: 79–90.
- Tensen, C.P., Van Kesteren, E.R., Planta, R.J., Cox, K.J., Burke, J.F., van Heerikhuizen, H., and Vreugdenhil, E. 1994. A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction. *Proc. Natl. Acad. Sci.* **91**: 4816–4820.
- Terhzaz, S., O'Connell, F.C., Pollock, V.P., Kean, L., Davies, S.A., Veenstra, J.A., and Dow, J.A. 1999. Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*. *J. Exp. Biol.* **202**: 3667–3676.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.

- Tissenbaum, H.A. and Ruvkun, G. 1998. An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans*. *Genetics* **148**: 703–717.
- Ui-Tei, K., Sakuma, M., Watanabe, Y., Miyake, T., and Miyata, Y. 1995. Chemical analysis of neurotransmitter candidates in clonal cell lines from *Drosophila* central nervous system, II: Neuropeptides and amino acids. *Neurosci. Lett.* **195**: 187–190.
- Vanden Broeck, J. 2001. Neuropeptides and their precursors in the fruitfly, *Drosophila melanogaster*. *Peptides* **22**: 241–254.
- Veenstra, J.A. 1994. Isolation and structure of the *Drosophila* corazonin gene. *Biochem. Biophys. Res. Commun.* **204**: 292–296.
- Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A., and Axel, R. 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**: 725–736.
- Williamson, M., Lenz, C., Winther, M.E., Nässel, D.R., and Grimmelikhuijzen, C. 2001. Molecular cloning, genomic organization, and expression of a B-type (cricket type) allatostatin preprohormone from *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **281**: 544–550.
- Xu, W.H., Sato, Y., Ikeda, M., and Yamashita, O. 1995. Molecular characterization of the gene encoding the precursor protein of diapause hormone and pheromone biosynthesis activating neuropeptide (DH-PBAN) of the silkworm, *Bombyx mori* and its distribution in some insects. *Biochim. Biophys. Acta* **1261**: 83–89.
- Yoon, J.G. and Stay, B. 1995. Immunocytochemical localization of *Diploptera punctata* allatostatin-like peptide in *Drosophila melanogaster*. *J. Comp. Neurol.* **363**: 475–488.
- Zdarek, J., Nachman, R.J., and Hayes T. 1997. Insect neuropeptides of the pyrokinin/PBAN family accelerate pupariation in the fleshfly (*Sarcophaga bullataria*) larvae. *Ann. NY Acad. Sci.* **814**: 67–71.
- Zhong, Y. and Pena, L.A. 1995. A novel synaptic transmission mediated by a PACAP-like neuropeptide in *Drosophila*. *Neuron* **14**: 527–536.
- Žitňan, D., Sehnal, F., and Bryant, P.J. 1993. Neurons producing specific neuropeptides in the central nervous system of normal and pupariation-delayed *Drosophila*. *Dev. Biol.* **156**: 117–135.

Received November 3, 2000; accepted in revised form March 22, 2001.