Biochemical and Biophysical Research Communications **269,** 91–96 (2000)
doi:10.1006/bbrc.2000.2251, available online at http://www.idealibrary.com on IDEAL®

Molecular Cloning and Genomic Organization of a Novel Receptor from *Drosophila melanogaster* Structurally Related to Mammalian Galanin Receptors[1]

Camilla Lenz,* Leif Søndergaard,† and Cornelis J. P. Grimmelikhuijzen*,[2]

**Department of Cell Biology, Zoological Institute, and †Department of Genetics, Institute of Molecular Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark*

Received January 7, 2000

We screened the Berkeley "*Drosophila* Genome Project" database with "electronic probes" corresponding to conserved amino acid sequences from the five known rat somatostatin receptors. This yielded alignment with a *Drosophila* genomic clone that contained a DNA sequence coding for a protein, having amino acid sequence identities with the rat galanin receptors. Using PCR with *Drosophila* cDNA as a template, and oligonucleotide probes coding for the exons of the presumed *Drosophila* gene, we were able to clone the cDNA for this receptor. The *Drosophila* receptor has most amino acid sequence identity with the three mammalian galanin receptors (37% identity with the rat galanin receptor type-1, 32% identity with type-2, and 29% identity with type-3). Less sequence identity exists with the mammalian opioid/nociceptin-orphanin FQ receptors (26% identity with the rat μ opioid receptor), and mammalian somatostatin receptors (25% identity with the rat somatostatin receptor type-2). The novel *Drosophila* receptor gene contains ten introns and eleven exons and is located at the distal end of the X chromosome. © 2000 Academic Press

Insects constitute 75% of all animals and are economically and ecologically extremely important because 70% of all flowering plants depend on insects for their pollination and insects can be severe pests, destroying about 30% of our potential annual harvest. Despite the importance of insects, however, the molecular endocrinological basis of central processes such as reproduction and development is still not well understood. This will certainly change by the forthcoming publication of the complete Berkeley "*Drosophila* Genome Project" database in the course of the year 2000 (1). This database, namely, will enable us to find novel neurohormone receptors and the corresponding ligands, and subsequently to elucidate the functions of these novel receptor/ligand couples. Because the complete database contains the information of *all* neurohormone receptors and preprohormones present in *Drosophila*, we will be able to determine *all* neurohormone receptor/ligand couples in an insect and, consequently, get a fully new insight into the endocrinology of insects.

So far, 25% of the *Drosophila* genome has been sequenced and published by the Berkeley *Drosophila* Genome Project. In the present paper, we will give an example of how to find novel neurohormone receptors in this partly completed database.

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. Poly(A)$^+$ RNA was prepared from third instar larvae of *D. melanogaster*, and single-stranded cDNA was prepared as in (2). PCR was performed as in (2), using the gene-specific sense primer 5′-ATGCGCTCCACCAC-CAATCTG-3′ (corresponding to nucleotide positions 313–333 of Fig. 2) and antisense primer 5′-GCGAAAGTTGTCGGATAGAAA-3′ (corresponding to nucleotide positions 1045–1065 of Fig. 2). The PCR program was 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min. 5′-RACE was carried out as a nested PCR with an antisense primer corresponding to nucleotide positions 654–678 and an antisense nested primer corresponding to nucleotide positions 635–655 of Fig. 2. For 3′-RACE, a sense primer corresponding to positions 654–678 and a nested sense primer corresponding to positions 813–833 of Fig. 2 was used. For the 5′- and 3′-RACEs, we used the SMART RACE cDNA Amplification method (Clontech). All PCR products were cloned into the *Eco*RI site of pCR-II-TOPO (Invitrogen) and sequenced by the use of the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing method (Amersham). Nucleotide and amino acid sequence comparisons and database searches were performed using the Lasergene software package (DNASTAR Inc).

[1] The nucleotide sequence reported in this paper has been submitted to the GenBank Data Bank with Accession No. AF220216.

[2] To whom correspondence should be addressed. Fax: 0045 35 32 12 00. E-mail: cgrimmelikhuijzen@zi.ku.dk. Homepage: www.zi.ku.dk/cellbiology.

0006-291X/00 $35.00

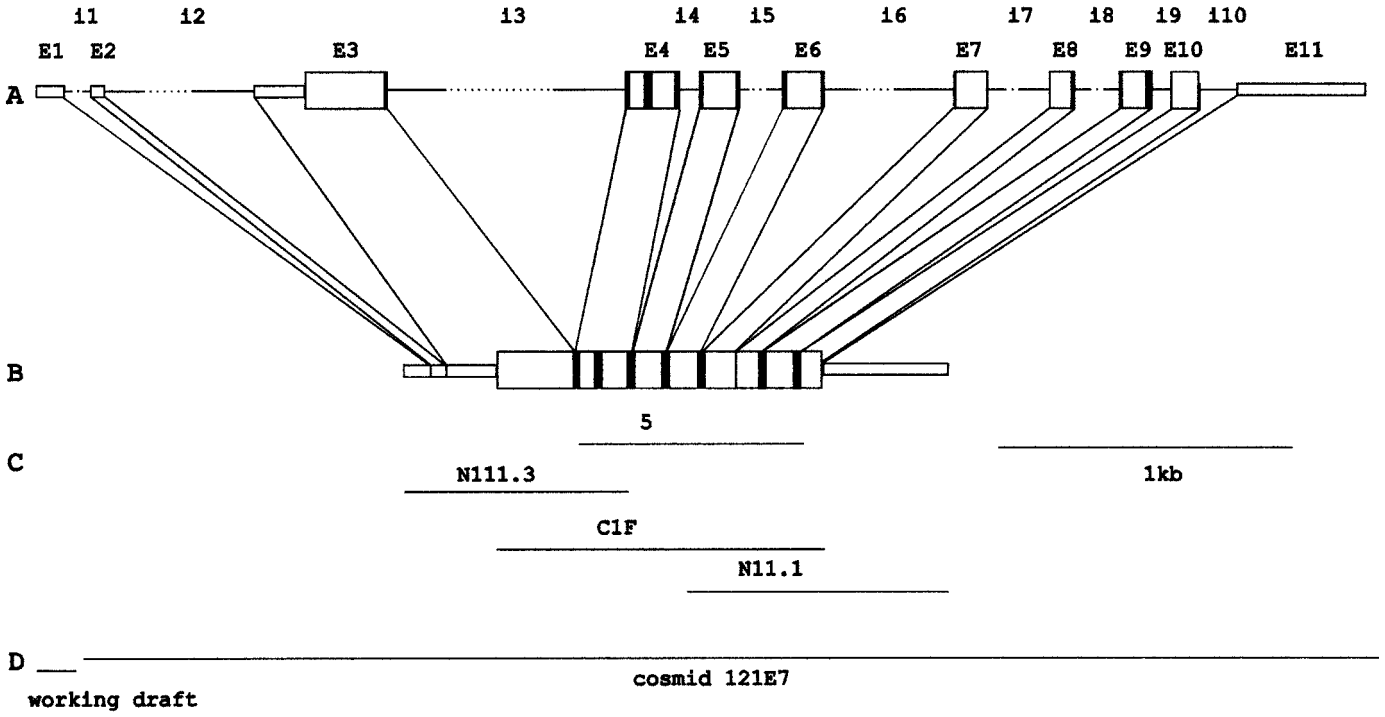

FIG. 1. Schematic representation of the *Drosophila* receptor gene, its cDNA, the locations of the various cDNA clones, and the locations of the two genomic clones. (A) Representation of the *Drosophila* receptor gene. The introns are given as lines and marked i1–i10. The exons are given as bars, being broad for the coding regions and small for the noncoding regions, and are marked E1–E11. The regions coding for the seven transmambrane helices are highlighted in gray. (B) The cDNA coding for the *Drosophila* receptor. (C) Positions of the various cDNA clones yielding the composite cDNA. The upper clone was the initial clone obtained by PCR, the two middle clones were obtained by 5′-RACE, and the lower clone was obtained by 3′-RACE. (D) The positions of the two genomic database clones, cosmid 121E7 (GenBank Accession No. AL 024454), and "working draft" (GenBank Accession No. AC 012898).

RESULTS

Cloning of the Drosophila receptor. We aligned the five known rat somatostatin receptors and obtained several regions with conserved (common) amino acid sequences. Using these sequences as "electronic probes", we screened all the available (July 1999) genomic sequences from the Berkeley *Drosophila* Genome Project database. This resulted in a significant alignment with DNA sequences from cosmid 121E7 (GenBank Accession No. AL024454), which already had been identified by collaborators of the Berkeley project as carring a putative gene. Using oligonucleotide probes against the presumed exons of this gene and *Drosophila* cDNA as a template, we could subsequently amplify the corresponding cDNA fragments by PCR (Fig. 1C), showing that the gene was indeed expressed. Using 3′- and 5′-RACE we could finally determine the 3′- and 5′-ends of this cDNA (Fig. 1C).

The complete, composite cDNA is shown in Fig. 2. This cDNA has a coding region of 1182 nucleotides and a 3′ untranslated region of 423 nucleotides, containing two polyadenylylation signals and the beginning of a poly(A^+) tail. The 5′ untranslated region is 340 bp long. Several 5′-RACE experiments were carried out, but did not give a longer 5′ untranslated region than the one shown in Fig. 2, suggesting that position −340 is the transcription start site.

The cDNA codes for a protein of 394 amino acid residues long (Fig. 2). The ATG codon at positions 1–3 is probably the translation start, because it is preceded by two in-frame stop codons (at positions −189 to −187 and −77 to −75 of Fig. 2). The protein has a signal sequence for RER membrane translocation that is probably cleaved off between Gly-24 and Ala-25 (3, 4). Hydropathicity plots show that the protein has seven transmembrane domains characteristic for G protein-

FIG. 2. cDNA and deduced amino acid sequence of the *Drosophila* receptor. The cDNA is composed of four clones (Fig. 1C). 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. Introns are indicated by arrows and numbered 1–10. The exon nucleotides bordering the introns are highlighted in gray. The seven membrane spanning domains are boxed and labeled TM I–VII. The translation termination codon is indicated by an asterisk. In-frame stop codons in the 5′-noncoding region are underlined. Putative polyadenylylation sites in the 3′-noncoding region are underlined twice. Putative glycosylation sites in the extracellular N terminus of the protein having the N-X-S or N-X-T consensus sequence are indicated by triangles.

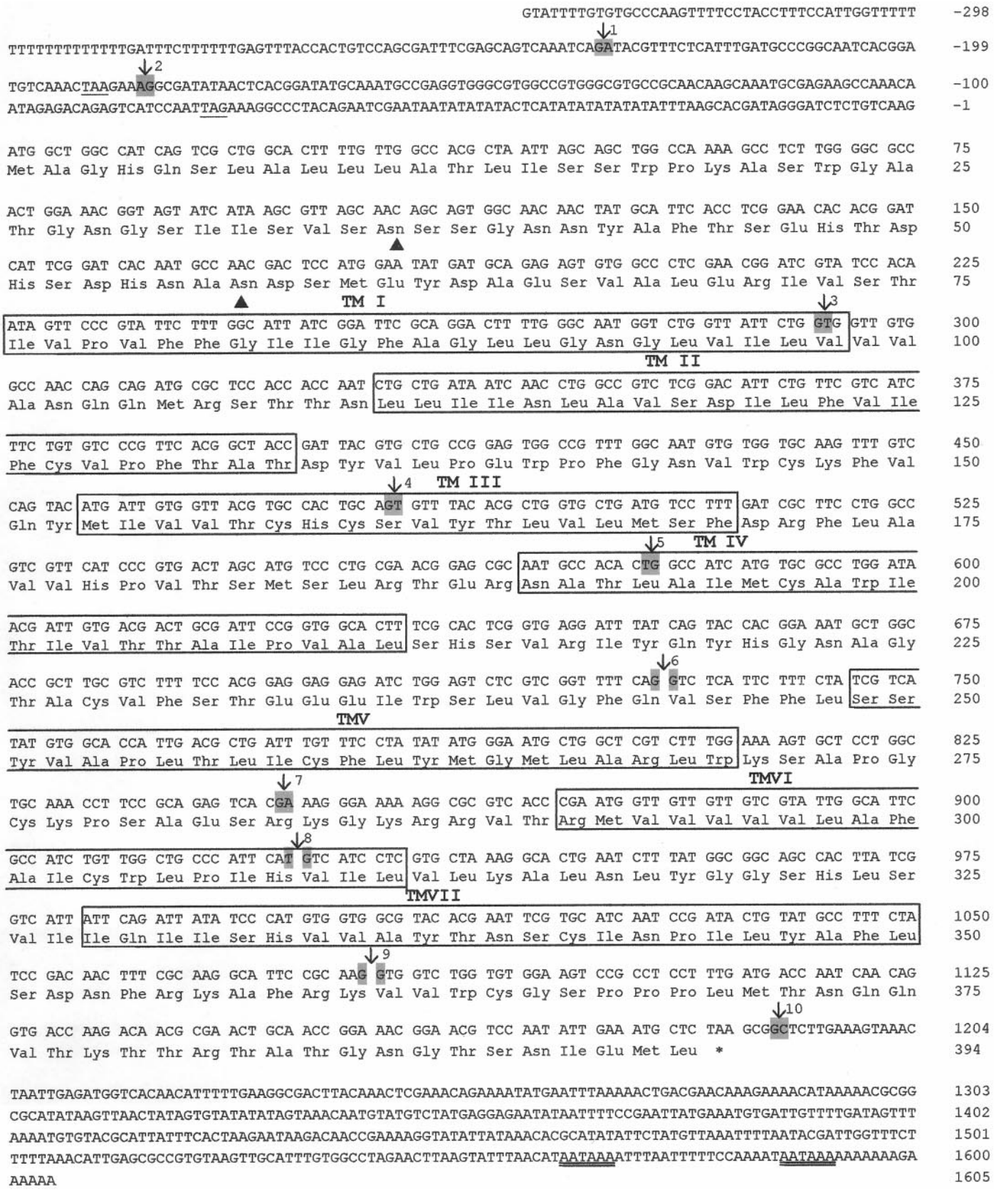
TM I
TM II
TM III
TM IV
TMV
TMVI
TMVII

FIG. 3. Amino acid sequence comparison of the *Drosophila* receptor (DGR) and the three rat galanin receptors (RGR-1, RGR-2, RGR-3). Amino acid residues that are identical between the *Drosophila* receptor and at least one of the other receptors are boxed. The seven membrane spanning domains are indicated by I–VII. Dashed lines represent spaces introduced to optimize alignment. The positions of the amino acid residues are given at the right. The data from the rat receptors are from (18–22).

coupled receptors. The extracellular N terminus has two potential glycosylation sites, whereas the intracellular C terminus has several Ser and Thr residues that are potential phosphorylation sites.

Comparison of the Drosophila receptor with other related receptors. Database searches show that the *Drosophila* protein has the highest amino acid sequence identity with the mammalian galanin receptors (37% amino acid identity with rat galanin receptor type-1; 32% with type-2; 29% with type-3; Fig. 3). Less amino acid sequence identities are found with the mammalian opioid/nociceptin-orphanin FQ receptors (maximally 26% identity with the rat μ opiod receptors), and somatostatin receptors (maximally 25% identity with rat somatostatin type-2 receptors).

Genomic organization of the Drosophila receptor. Alignment of the *Drosophila* receptor cDNA with the genomic DNA sequence of cosmid 121E7 and with a recently published "working draft" (GenBank Accession Nos. AL024454 and AC012898), shows that the receptor gene contains 10 introns and 11 exons (Figs. 1A and 1B; Table 1). Seven of these introns are found in the region coding for the seven-transmembrane domain of the receptor, which is rather unusual for G protein-coupled receptor genes.

There is a small number of nucleotide differences between our cloned cDNA and the genomic DNA sequences of cosmid 121E7 and the "working draft" (Table 2). These nucleotide differences, however, do not lead to differences in amino acid residues.

TABLE 1

Intron/Exon Boundaries of the *D. melanogaster* Gene

Intron	5′-Donor	Intron size (bp)	3′ Acceptor	Intron phase
1	CAG gtgagtt...	>4268	...attgcag ATA	—
2	AAA gtgagtg...	13236	...ctttcag GGC	—
3	G gtaagtg...	14575	...gttgcag TG Leu	1
4	AG gtgggtg... Ser	81	...cttgcag T Ser	2
5	CT gtgagtg... Leu	722	...tccttag G Leu	2
6	CAG gttagtt... Gln	5342	...atttcag GTC Val	3
7	CG gtaagta... Arg	1146	...atttcag A Arg	2
8	CAT gtgagta... His	802	...tttgcag GTC Val	3
9	AAG gtgggca... Lys	71	...cttgcag GTG Val	3
10	CGG gtatgta...	141	...cttgcag CTC	—

Note. The sequence of each of the intron/exon bounderies is shown, as well as the codons for the amino acid residues. Uppercase and lowercase letters represent nucleotides in the exons and introns, respectively. The sequence of the introns can be retrieved from the Berkeley *Drosophila* Genome Project database, Accession Numbers AL024454 (cosmid 121E7) and AC012898 ("working draft"). Because the two clones from the Berkeley database do not overlap, the size of intron 1 is unknown, but it must be larger than 4268 bp. The overall positions of the introns are shown in Figs. 1 and 2.

Chromosomal localization of the Drosophila receptor. Based on the chromosomal localization of cosmid 121E7 by the Berkeley *Drosophila* Genome Project, our *Drosophila* receptor gene is localized on the distal end of the X chromosome.

DISCUSSION

We have mined the Berkeley *Drosophila* Genome Project database with an "electronic probe" based on conserved regions from the five known rat somatostatin receptors. This resulted in the alignment of the probe with a DNA sequence, coding for a novel putative *Drosophila* G protein-coupled receptor. We subsequently cloned the complete cDNA coding for this receptor, showing that the putative *Drosophila* gene was indeed expressed. GenBank database alignments showed that the novel *Drosophila* receptor was most closely related to the mammalian galanin receptors and to a lesser extend to the mammalian opioid and somatostatin receptors. Our experiments, therefore, show that it is feasible to screen the *Drosophila* Genome Project database with "electronic probes" based on conserved regions from family members of mammalian neurohormone (G protein-coupled) receptors. This approach will ultimately enable us to find most or perhaps all neurohormone receptors in insects (when the *Drosophila* Genome Project is completed) and, subsequently, to isolate the corresponding ligands. These developments will revolutionize our knowledge of the insect neuroendocrine system and will lead to a far better understanding of insects than has hitherto been possible.

Shortly after our cDNA cloning had been completed (Fig. 2), another research group very recently published a similar cDNA sequence (5). This sequence had not been obtained by the database mining approach described above, but by "classical" PCR, using oligonucleotide probes corresponding to rat somatostatin receptors. The cDNA sequence from the other group (5) is similar to ours, but our sequence contains more 5′-cDNA information, including the transcription start site, which was not determined by the other group. Furthermore, we established the genomic organization of the novel receptor gene (Figs. 1A, 2), which was not fully correctly determined by the other research group (5).

In addition to the above described receptor, 12 neurohormone (G protein-coupled) receptors have been cloned from *Drosophila*. These include three different receptors for serotonin (6, 7), two different receptors for dopamine (8, 9), one receptor for octopamine (10, 11), acetylcholine (12), the tachykinins (13, 14), and NPY (15), two orphan receptors related to mammalian glycoprotein hormone receptors (16; Eriksen *et al.*, submitted], and one orphan receptor related to vertebrate GnRH receptors (17). It is difficult to say, how much this number will increase, when the *Drosophila* Genome Project database will be completed and unravelled. Based on the situation in other animal groups,

TABLE 2

Nucleotide Differences between the cDNA of Fig. 2 and the Corresponding Genomic Sequences from the Berkeley *Drosophila* Genome Project

Position of the nucleotide in the cDNA	Type of nucleotide(s) in the gene	Type of nucleotide in the cDNA	Change in amino acid
−334	G	T	—
between −284 and −285	TT	absent	—
498	C	G	Val → Val
1273	absent	A	—
1277	absent	A	—
1278	absent	C	—
1279	absent	G	—
1280	absent	A	—

Note. The position of the nucleotide in the cDNA (Fig. 2) is given in the first column, the type of nucleotide present in the genomic sequence in the second column, and the type of nucleotide present in the cDNA in the third column. Most changes lie outside the coding region, one change lies within the coding region, but here the changed nucleotide does not result in a changed amino acid residue.

e.g. mammals (at least 200–300 neurohormone receptors present), however, it is safe to say that the total number of neurohormone receptors in an insect will be over 100. The next coming years, therefore, are exciting years for insect molecular neuroendocrinologists.

ACKNOWLEDGMENTS

We thank Lotte Steffensen for typing the manuscript, and the Danish Biotechnological Research and Development Program of the Danish Research Councils, and Novo Nordisk Foundation for financial support.

REFERENCES

1. Butler, D. (1999) *Nature* **401,** 729–730.
2. Hauser, F., Williamson, M., and Grimmelikhuijzen, C. J. P. (1997) *Biochem. Biophys. Res. Commun.* **241,** 509–512.
3. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
4. Nielsen, H., Engelbrecht, J., Brunak, S., and Von Heijne, G. (1997) *Protein Eng.* **10,** 1–6.
5. Birgül, N., Weise, C., Kreienkamp, H.-J., and Richter, D. (1999) *EMBO J.* **18,** 5892–5900.
6. Witz, P., Amlaiky, N., Plassat, J.-L., Maroteaux, L., Borrelli, E., and Hen, R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8940–8944.
7. Saudou, F., Boschert, U., Amlaiky, N., Plassat, J. L., and Hen, R. (1992) *EMBO J.* **11,** 7–17.
8. Gotzes, F., Balfanz, S., and Baumann, A., (1994) *Receptors Channels* **2,** 131–141.
9. Feng, G., Hannan, F., Reale, V., Hon, Y. Y., Kousky, C. T., Evans, P. D., and Hall, L. M. (1996) *J. Neurosci.* **16,** 3925–3933.
10. Arakawa, S., Gocayne, J. D., McCombie, W. R., Urquhart, D. A., Hall, L. M., Fraser, C. M., and Venter, J. C. (1990) *Neuron* **4,** 343–354.
11. Saudou, F., Amlaiky, N., Plassat, J. L., Borrelli, E., and Hen, R. (1990) *EMBO J.* **9,** 3611–3617.
12. Shapiro, R. A., Wakimato, B. T., Subers, E. M., and Nathanson, N. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9039–9043.
13. Li, X. J., Wolfgang, W., Wu, Y. N., North, R. A., and Forte, M. (1991) *EMBO J.* **10,** 3221–3229.
14. Monnier, D., Colas, J.-F., Rosay, P., Hen, R., Borrelli, E., and Maroteaux, L. (1992) *J. Biol. Chem.* **267,** 1298–1302.
15. Li, X. J., Wu, Y. N., North, R. A., and Forte, M. (1992) *J. Biol. Chem.* **267,** 9–12.
16. Hauser, F., Nothacker, H.-P., and Grimmelikhuijzen C. J. P. (1997) *J. Biol. Chem.* **272,** 1002–1010.
17. Hauser, F., Søndergaard, L., and Grimmelikhuijzen, C. J. P. (1998) *Biochem. Biophys. Res. Commun.* **249,** 822–828.
18. Parker, E. M., Izzarelli, D. G., Nowak, H. P., Mahle, C. D. Iben, L. G. Wang, J., and Goldstein, M. E. (1995) *Mol. Brain Res.* **34,** 179–189.
19. Howard, A. D., Tan, C., Shiao, L. L., Palyha, O. C., McKee, K. K., Weinberg, D. H., Feighner, S. D., Cascieri, M. A., Smith, R. G., Van der Ploeg, L. H., and Sullivan, K. A. (1997) *FEBS Lett.* **405,** 285–290.
20. Wang, S., Hashemi, T., He, C., Strader, C., and Bayne, M. (1997) *Mol. Pharmacol.* **52,** 337–343.
21. Smith, K. E., Walker, M. W., Artymyshyn, R., Bard, J., Borowsky, B., Tamm, J. A., Yao, W. J., Vaysse, P. J.-J., Branchek, T. A., Gerald, C., and Jones, K. A. (1998) *J. Biol. Chem.* **273,** 23321–23326.
22. Wang, S., He, C., Hashemi, T., and Bayne, M. (1997) *J. Biol. Chem.* **272,** 31949–31952.