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*

(Received for publication, February 22, 1996, and in revised form, October 16, 1996)

Frank Hauser‡, Hans-Peter Nothacker§, and Cornelis J. P. Grimmelikhuijzen¶

From the Department of Cell Biology and Anatomy, Zoological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Using oligonucleotide probes derived from consensus sequences for glycoprotein hormone receptors, we have cloned an 831-amino acid residue-long receptor from Drosophila melanogaster that shows a striking structural homology with members of the glycoprotein hormone (thyroid-stimulating hormone (TSH); follicle-stimulating hormone (FSH); luteinizing hormone/choriogonadotropin (LH/CG)) receptor family from mammals. This homology includes a very large, extracellular N terminus (26% sequence identity with rat TSH, 19% with rat FSH, and 20% with the rat LH/CG receptor) and a seven-transmembrane region (53% sequence identity with rat TSH, 50% with rat FSH, and 52% with the rat LH/CG receptor). The Drosophila receptor gene is >7.5 kilobase pairs long and contains 17 exons and 16 introns. Seven intron positions coincide with introns in the mammalian glycoprotein hormone receptor genes and have the same intron phasing. This indicates that the Drosophila receptor is evolutionarily related to the mammalian receptors. The Drosophila receptor gene is located at position 90C on the right arm of the third chromosome. The receptor is strongly expressed starting 8–16 h after oviposition, and the expression stays high until after pupation. Adult male flies express high levels of receptor mRNA, but female flies express about 6 times less. The expression pattern in embryos and larvae suggests that the receptor is involved in insect development. This is the first report on the molecular cloning of a glycoprotein hormone receptor family member from insects.

Insects are very successful and adaptive animals and are of extreme, ecological and economical importance, since about 70% of all flowering plants depend on insects for pollination and insects can be severe pest animals. Despite the importance of insects, however, the molecular basis of their reproduction is not understood very well. This is in contrast to the situation in mammals, where reproduction is controlled by (i) steroid hormones, (ii) a family of closely related glycoprotein hormones produced by the adenohypophysis (luteinizing hormone (LH)1; follicle-stimulating hormone (FSH)), or placenta (choriogonadotropin (CG)), and (iii) hypothalamic releasing hormones (e.g. gonadotropin-releasing hormone) that control the release of LH and FSH from the adenohypophysis. A few years ago, the mammalian receptors for FSH, LH, and CG were cloned, as well as the receptor for a fourth glycoprotein hormone, thyroid-stimulating hormone (TSH) that also is produced by the adenohypophysis (1–9). The receptors for human LH and CG appeared to be identical and closely related to the TSH and FSH receptors (4–8, 10). All three receptor types form a distinct subfamily belonging to the large family of G protein-coupled (seven-transmembrane) receptors. A remarkable feature of the mammalian TSH, FSH, LH/CG receptors, which separate them from the other G protein-coupled receptors, is the presence of a very large, extracellular N terminus, which composes about half of the receptor molecule. This N terminus probably constitutes the binding site for the glycoprotein hormone ligand (11–13).

Recently, using the polymerase chain reaction (PCR) and primers coding for consensus sequences of G protein-coupled receptors, we have found that cnidarians (which are the most primitive animals in the animal kingdom having a nervous system, such as sea anemones) produce a receptor that shows a striking structural homology with members of the mammalian TSH, FSH, LH/CG receptors (14). This was an exciting finding, since glycoprotein hormone receptors had, so far, only been cloned from mammals and not from other vertebrates or invertebrates. Our results suggested that glycoprotein hormone receptors and possibly all of the processes that are mediated by these types of receptors are conserved throughout the animal kingdom, from cnidarians to mammals. The presence of putative glycoprotein hormone receptors in invertebrates was confirmed shortly after our own discovery in cnidarians by a report on the cloning of a related receptor in molluscs (15). In the near future, it will be interesting to see if these receptors are similar to the mammalian receptors, if they are expressed in the same manner and if they mediate similar processes.

* This work was supported in part by the Danish Biotechnological Research and Development Program of the Danish Research Councils, Carlsberg Foundation, Løg- og Naturvidenskabelige Komité of the Novo Nordisk Foundation, and Fonds der Chemischen Industrie (Germany). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank/EBI Data Bank with accession number(s) U47005, U47006, and U49733.

‡ Recipient of a postdoctoral fellowship from the Max Planck Society (Germany).
§ Present address: CNS Research, Pharma Devisjon, P. Hoffmann-La Roche AG, CH-4070 Basel, Switzerland.
¶ To whom correspondence should be addressed: Dept. of Cell Biology and Anatomy, Zoological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark. Tel.: 0045-35 32 12 27; Fax: 0045-35 32 12 00.

1 The abbreviations used are: LH, luteinizing hormone; bp, base pairs; CG, choriogonadotropin; FSH, follicle-stimulating hormone; kb, kilobase pairs; PCR, polymerase chain reaction; RP49, ribosomal protein 49; TSH, thyroid-stimulating hormone; RACE, rapid amplification of cDNA ends.
present paper, we have focused on insects, and we have cloned and characterized a *Drosophila melanogaster* receptor that shows a strong, structural homology with the TSH, FSH, LH/CG receptors from mammals.

**EXPERIMENTAL PROCEDURES**

**Radioactive Labeling of DNA Probes**—DNA fragments to be labeled were cleaved from vector DNA by restriction enzymes and purified by agarose gel electrophoresis. They were labeled with $\alpha$-$^{32}$PdCTP (Amersham Corp.; specific activity, 110 TBq/mmol) using the Ready-To-Go DNA-labelling kit from Pharmacia Biotech. Inc. (16). DNA Sequencing and Sequence Analysis—DNA sequences were determined by the dideoxynucleotide chain termination method (17) using a Sequenase version 2.0 DNA sequencing kit (U.S. Biochemical Corp.). cDNA insertions were excised from the lambda vector with restriction enzymes, subcloned into pBluescript SK (Stratagene), and sequenced. All subclones were sequenced in both directions. For the sequencing of larger insertions, the Double Stranded Nested Deletion kit (Pharmacia) or sequence-specific oligonucleotides were used. DNA sequence compilations, nucleotide and amino acid sequence comparisons, and data base searches were performed using the Lasergene software package (DNASTAR Inc.).

**PCR**—In order to obtain the initial PCR product to be used to screen the cDNA library, we used two primers, sense 5'-CCITG(TG/CA/AA/GGA/G/AGA/TATAATGTTGITA-3' and antisense 5' -GC(G/A)AAAGGCAC/G/AGA/GA/TTATAGCACATIAG/G/AA-3' and a AZAP1 cDNA library from adult *D. melanogaster* (a kind gift from Dr. E. Gundelfinger, Federal Institute for Neurobiology, Magdeburg, Germany) as a template to perform PCR (18). The reaction mixture consisted of 50 μl of 1 × Taq buffer (Promega) containing 2.5 mM MgCl₂, 1 × 10⁶ plaque-forming units of cDNA library, 0.8 μM of each primer, 0.2 mM of all four dNTPs, and 1.5 units of Taq polymerase (Promega). Thermal cycling parameters were 3 min of initial denaturation at 94°C, followed by 30 cycles of the following step program: 94°C for 1 min, 45°C for 1 min, 72°C for 1 min. The PCR products were separated on 2% agarose gel, and bands of the expected size were isolated (Qiaquick extraction kit, Qiagen), subcloned into the Smal site of PCR-Script (Stratagene), and sequenced.

**5'-RACE**—All reactions of the 5'-RACE were carried out following the protocol and using the chemicals of the 5'/3'-RACE kit of Boehringer Mannheim. mRNA from *D. melanogaster* Canton S embryos (16–24 h after oviposition) was used as template. mRNA was isolated with the Oligotex Direct mRNA kit (Qiagen). We used an antisense primer corresponding to positions 115–135 of Fig. 2 for cDNA synthesis and two nested antisense primers corresponding to positions 40 to 63 and 175 to 146 (Fig. 2) for the extension reaction. The PCR products were obtained with blunt ends using Taq DNA polymerase (Pharmacia), phosphorylated using T4 polynucleotide kinase (Amersham) (19), subcloned into the Smal site of pBluescript SK (Stratagene), and sequenced.

**Primer Extension**—Primer extension was carried out according to Ref. 19 using mRNA from embryos (16–24 h after oviposition) purified as described above. We used an antisense primer corresponding to positions 174 to 146 of Fig. 2 for the extension reaction.

**Isolation of cDNA Clones**—The cDNA insertion of one of the PCR clones (clone pBD) was $^{32}$P-labeled and used as a probe to screen a Agt11 cDNA library from adult *D. melanogaster* Canton S (Clontech). A second cDNA library from adult *D. melanogaster* Canton S, which was cloned into the BamHI and XbaI sites of ADR2 (Clontech), was screened with a $^{32}$P-labeled insertion of clone dgr1. Plaque lifting and processing of the nitrocellulose filters were performed as described in Ref. 19.

**Isolation of Genomic Clones**—A genomic AFl II library from embryos of *D. melanogaster* Canton S (Stratagene) was screened with a $^{32}$P-labeled insertion of cDNA clone dgr1. Further procedures were as described above.

**Northern Blot Analysis**—mRNA from 10² to 10⁴ animals was isolated with the Oligotex Direct mRNA kit (Qiagen), electrophoresed on 1% agarose, 6% formaldehyde gels, and capillary-transferred onto Hybond-N membranes (Amersham) as described (19). After baking for 1 h at 80°C, the blot was prehybridized for 4 h at 42°C in a solution containing 50% (v/v) formamide, 5 × SSC (1 × SSC: 150 mM NaCl, 15 mM sodium citrate, pH 7.0), 5 × Denhardt’s solution (1 × Denhardt’s solution: 0.7 M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1% SDS, 0.1 mg/ml each of yeast tRNA, salmon sperm DNA, and poly-dI-dC).
FIG. 2. cDNA and deduced amino acid sequence of the *Drosophila* receptor. This figure consists of the insertions of cDNA clone *ldg8* (positions 2254 to 2925) and of the longest cDNA clone derived by 5'-RACE PCR (positions 2265 to 2414). Nucleotides are numbered from 5'- to 3'-end, and the amino acid residues are numbered starting with the first ATG in the open reading frame. Introns are indicated by arrows and numbered 1–16. The seven membrane-spanning domains are boxed and labeled TM I–VII.
solution: 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll), 0.5% SDS, 0.002% herring sperm DNA. For hybridization, the heat-denatured radioactive probe was added at a concentration of 2 × 10^6 cpm/ml and incubated for 18 h at 42°C. Final washing was 2 × 10 min at 50°C in 0.1 × SSC, 0.1% SDS. To obtain a hybridization probe for the ribosomal protein 49 (RP49) mRNA, positions −347 to +689 of the RP49 gene (20) were PCR-amplified using two primers, both 23 nucleotides long, and genomic DNA from Drosophila as a template. The hybridization signals of the Drosophila receptor and RP49 mRNAs were quantified by scanning of the autoradiograms and integration using the CREAM™ software package (Korn-En-Tec A/S, Copenhagen). As a size marker, the DNA Molecular Weight Marker II (Boehringer Mannheim) was used.

Southern Blot Analysis—D. melanogaster Canton S genomic DNA (Clontech) was digested with various restriction enzymes and separated in 0.8% agarose gels. Transfer to a Hybond-N membrane, hybridization, and washing were done as recommended by the supplier (Amersham).

Receptor Gene Localization on Polytenic Chromosomes—D. melanogaster salivary gland polytene chromosome squashes were prepared on microscopic slides as described (21). These slides were incubated in 2 × SSC for 30 min at 65°C, after which the DNA was denatured for 2 min in 0.07 M NaOH. After neutralization in 2 × SSC, the slides with the chromosome squashes were washed with increasing concentrations of ethanol and air-dried. The preparations were subsequently hybridized at 58°C for 16 h in a solution containing 5 × SSC, 0.1% SDS, and a heat-denatured digoxigenin-labeled cDNA probe. The labeling was performed with a digoxigenin DNA-labeling kit (Boehringer Mannheim). Washing was for 30 min at 53°C in 2 × SSC. Sites of hybridization were visualized using digoxigenin antibodies conjugated with horseradish peroxidase and diaminobenzidine/H2O2 as substrate. Chromosomes were stained with Giemsa stain (Fluka).

RESULTS
cDNA Cloning of a Drosophila Receptor Protein—We constructed one sense PCR primer that was based on conserved amino acid sequences located N-terminally of the first membrane-spanning domain of several mammalian glycoprotein hormone receptors and a second antisense primer corresponding to conserved amino acid sequences of the second membrane-spanning region. Using these primers and a cDNA library from adult D. melanogaster as a template, we carried out PCR. After subcloning of the PCR products, we obtained one clone (pD5) that contained a 161-bp insertion coding for a protein fragment comprising the first and second membrane-spanning regions of a putative Drosophila receptor related to the TSH, FSH, and LH/CG receptors from mammals. Subsequently, we screened 5 × 10^5 phages of a cDNA library of adult D. melanogaster Canton S with the cDNA insertion of clone pD5 and obtained three positive clones (λdgr1, λdgr2, and λdgr4). These clones contained cDNA insertions of 2520, 2159, and 1248 bp, respectively, all coding for the same transcript (Fig. 1A). All three insertions terminated at their 3′-ends with the same internal EcoRI restriction site (Fig. 1A). We therefore used the insertion of clone λdgr1 to screen 5 × 10^5 phages of another D. melanogaster Canton S cDNA library, which was subcloned into the BamHI and XhoI sites of λDR2. This resulted in the isolation of two positive clones, of which one (clone λdgr8) contained the longest cDNA (Figs. 1 and 2).

The transcription start site was determined using 5′-RACE. This yielded a major start site at position −264 of Fig. 2, which added 11 nucleotides to the 5′-end of clone λdgr8. This major transcription start site was confirmed by primer extension. Other 5′-RACE products pointed to additional transcription start sites at positions −261 and −246.

The composite cDNA of Fig. 2 is 3190 bp long. The ATG codon at nucleotide positions 1–3 of Fig. 2 is probably the start codon, because it is flanked by a consensus sequence for the initiation of translation (22). Furthermore, there is an in-frame stop codon upstream of this start codon (at positions −33 to −31). The cDNA codes for an 831-amino acid residue-long protein with a calculated molecular mass of 92.7 kDa. Hydropathy analysis suggests the presence of seven membrane-spanning regions, which is characteristic of G protein-coupled receptors.

The protein contains four potential N-glycosylation sites in the putative extracellular N terminus and two others in the extracellular loops of the transmembrane domain, suggesting that the actual molecular mass of the receptor might be higher than 92.7 kDa. The intracellular loops of the putative transmembrane domain and the C-terminal, intracellular part of the protein contain multiple serine and threonine residues, which are potential phosphorylation sites.

Comparison of the Drosophila Receptor with Other Related Receptors—Fig. 3 shows a comparison of the Drosophila receptor with the TSH, FSH, and LH/CG receptors from rat and the putative glycoprotein hormone receptor from sea anemones. Especially in the seven-transmembrane area (from the Drosophila receptor amino acid positions 493–759 in Fig. 3), there is a high sequence homology between the Drosophila and the mammalian receptor proteins (53% sequence identity with rat TSH, 50% with rat FSH, and 52% with rat LH/CG receptor). Less sequence identity is found in the extracellular N terminus (amino acid positions 1–492 in Fig. 3; 20% sequence identity with rat TSH, 19% with rat FSH, and 20% with rat LH/CG receptor).

One feature of the mammalian glycoprotein hormone receptors is the existence of Leu (or Ile/Val)-rich repeats in their N termini (12, 13, 23). These Leu-rich repeats also occur in the N terminus of the Drosophila receptor (Fig. 4). The positions of most of the aliphatic residues in the Leu-rich repeats of the Drosophila receptor follow a distinct pattern, which is very similar to that found in the mammalian receptors (Fig. 4; asterisks and filled circles in Fig. 3).

DNA Sequence and Organization of the Drosophila Receptor Gene—Using the cDNA insertion of clone λdgr1 as a probe, we screened 2.5 × 10^6 phages of a genomic library of D. melanogaster Canton S embryos and identified 21 positive clones. One positive clone (λg7/3), which contained the Drosophila receptor gene, was digested with restriction enzymes, and overlapping fragments were subcloned and sequenced (Fig. 1D). The DNA of the receptor gene is >7.5 kb long. It consists of 17 exons and 16 introns (Fig. 1C). Only introns 1, 2, and 5 are more than 400 bp in length. The others are considerably shorter, in the range of 30–150 bp (Fig. 1C; Table I). The donor and acceptor sites of the introns are in agreement with the GT/AG consensus sequence (24), with the exception of the 5′-donor site of intron 8, which contains an unusual GC sequence (Table I). The existence of this unusual GC sequence was confirmed by sequencing of another, independent genomic clone.

Intron 1, which has a size of 2.3 kb, is located in the 5′-noncoding region, whereas all of the other introns are located in the coding region of the gene (Fig. 1, B and C, and Fig. 2). Exon 1 (Fig. 1C) contains most of the 5′-noncoding region of the corresponding receptor mRNA. Exons 2–11 code for the extracellular, N-terminal domain of the receptor, exons 12–16 for the transmembrane regions, and exon 17 for the intracellular C-terminal part (Fig. 1, B and C, and Fig. 2). Exon 17 also contains the 3′-noncoding region.

The compiled DNA in the coding regions of the Drosophila receptor gene is identical to that of the cDNA sequence with a total of 16 exceptions (Table II). In these exceptions, the coding

_asterisk_: An in-frame stop codon in the 5′-noncoding region is underlined. A putative polyadenylylation site at the 3′-end is underlined twice. Putative glycosylation sites following the N-X-S or N-X-T consensus sequence are marked by triangles.
FIG. 3. Amino acid sequence comparison of the Drosophila receptor (GHRD), the rat TSH (TSHR), FSH (FSHR), and LH (LHR) receptors and the putative glycoprotein hormone receptor from the sea anemone Anthopleura elegantissima (GHRA). Amino acid residues that are identical between the Drosophila receptor and at least one of the other receptors are boxed. Known intron-exon transitions in the genes coding for four receptors are highlighted in gray at the corresponding amino acid residues. The intron-exon transitions in the rat TSH receptor gene are not known, but in other mammalian TSH receptor genes the first eight intron-exon transitions occur at exactly the same positions as in the genes for the rat FSH and LH receptor. The positions of the aliphatic residues characteristic for the Leu-rich repeats in the Drosophila receptor (see Fig. 4) are marked by asterisks or filled circles. The filled circles also mark intron-exon transitions in the Drosophila receptor gene that occur at the same positions and have the same intron phasing (see Table I) as in the mammalian receptor genes. The seven membrane-spanning domains are indicated by I–VII. Dashed lines represent spaces introduced to optimize alignment. The amino acid positions are given at the right. The amino acid sequences of the sea anemone and rat receptors and the intron-exon positions in their genes are from Refs. 3, 8, 14, 23, 28, and 29.
distinct exon (data from Ref. 23).

Thesegments L1–L7 each correspond to a
the N terminus of the rat FSH receptor.

alignment of Leu (Ile/Val)-rich repeats of

S strain from which the commercial genomic and cDNA libraries
may point to two subpopulations of the

The aliphatic residues highlighted in gray
correspond to intron-exon transitions in
the receptor gene and are given at the
start of seven repeating segment. The
segments L2–L7, therefore, each corre-
spond to a distinct exon. Amino acid positions
are given at the right. B, a similar
alignment of Leu (Ile/Val)-rich repeats of
the N terminus of the rat FSH receptor.

The segments L1–L7 each correspond to a
distinct exon (data from Ref. 23).

Intron/exon boundaries of the D. melanogaster receptor gene

The sequence of each of the intron-exon boundaries is shown, as well as
the codons for the amino acid residues. Uppercase and lowercase
letters represent nucleotides in the exons and introns, respectively.

Each intron has been fully sequenced with the exception of intron 1, of
which only 300 bases have been determined. The sequences of the
introns are given in our GenBank™/EMBL Data Bank submission.

The overall positions of the introns are shown in Figs. 1 and 2. Introns 3–9
occur at the same positions within the gene as seven introns in the

glycoprotein hormone receptor genes of mammals (cf. filled circles
in Fig. 3). The triplets coding for the aliphatic amino acid residues at
intron-exon transitions 9–9 are always interrupted after the second
nucleotide (intron phase = 2).

Intron 5

**Intron-exon boundaries of the D. melanogaster receptor gene**

The sequence of each of the intron-exon boundaries is shown, as well as
the codons for the amino acid residues. Uppercase and lowercase
letters represent nucleotides in the exons and introns, respectively.

Each intron has been fully sequenced with the exception of intron 1, of
which only 300 bases have been determined. The sequences of the
introns are given in our GenBank™/EMBL Data Bank submission.

The overall positions of the introns are shown in Figs. 1 and 2. Introns 3–9
occur at the same positions within the gene as seven introns in the

glycoprotein hormone receptor genes of mammals (cf. filled circles
in Fig. 3). The triplets coding for the aliphatic amino acid residues at
intron-exon transitions 9–9 are always interrupted after the second
nucleotide (intron phase = 2).

| Intron | 5’ Donor | Intron size (bp) | 3’ Acceptor | Intron phase |
|--------|----------|-----------------|-------------|-------------|
| 1      | CCG      | 2300            | GGG         | 1           |
| 2      | CAA      | 505             | ACC         | 2           |
| 3      | Gln      | 88              | Gly         | 3           |
| 4      | AT       | 80              | Val         | 4           |
| 5      | AT       | 62             | Ile         | 5           |
| 6      | AT       | 61             | Ile         | 6           |
| 7      | TT       | 63             | Ile         | 7           |
| 8      | CT       | 55             | Leu         | 8           |
| 9      | CT       | 67             | C9          | 9           |
| 10     | CGG      | 59             | AAT         | 10          |
| 11     | AG       | 63             | A2          | 11          |
| 12     | AG       | 144            | A3          | 12          |
| 13     | G        | 71             | A4          | 13          |
| 14     | AG       | 68             | C5          | 14          |
| 15     | TG       | 70             | C6          | 15          |
| 16     | AA       | 68             | A7          | 16          |

*TABLE I*

**Codon differences between the DNA and the gene coding for the Drosophila receptor**

Amino acid residue and position within the receptor protein (Fig. 2) is
given in the first column, the triplet in the gene in the second column,
and the triplet in the cDNA in the third column. None of the codon
differences leads to a difference in amino acid residue.

Amino acid residue and position | Codon in the gene | Codon in the cDNA |
---|---|---|
Gly374 | Gly | GGG |
Thr374 | Thr | ACT |
Phe495 | Phe | TCT |
Leu218 | Leu | CTC |
Leu227 | CTA |
Leu229 | CTA |
Ser138 | Ser | CCT |
Leu335 | TTG |
Arg118 | GAG |
Glu345 | GAG |
Val123 | GTC |
Leu134 | CTG |
Ala360 | CTT |
Gly366 | Gly | GGA |
Ser204 | Ser | AGT |
Pro273 | Pro | CCT |

*TABLE II*

triplets were always changed in such a way that no change
in the amino acid residues resulted (Table II). Eight nucleotide
differences occur between the genomic DNA and the 3’-noncoding
region of the receptor cDNA (given in our GenBank™
EMBL Data Bank submission). All of these small differences
may point to two subpopulations of the D. melanogaster Canton
S strain from which the commercial genomic and cDNA libraries
have been constructed, or to allelic variation.

The cDNA clone Adgr1 contains an additional 70-bp insertion

in the 5’-noncoding region compared with cDNA clone Adgr8
(Fig. 1A). This insertion probably originates from alternative
splicing at the 3’-end of the first intron (Fig. 5).

Seven intron-exon transitions in the DNA coding for the
extracellular N terminus of the Drosophila receptor occur at
exactly the same positions as in the genes for the mammalian
TSH, FSH, and LH/CG receptors (see those residues that are
marked by a filled circle in Fig. 3). This indicates that the
Drosophila receptor and the mammalian glycoprotein hormone
receptors are evolutionarily related.
We have cloned a novel \textit{Drosophila} receptor that shows a strong, structural homology with members of the TSH, FSH, LH/CG receptor family from mammals. This conclusion is based on the following findings. (i) The \textit{Drosophila} receptor is a seven-transmembrane receptor having a very large, extracellular N terminus that comprises about half of the total receptor molecule (Fig. 2). This overall structure is similar to that of the mammalian glycoprotein hormone receptors. (ii) There is 19–20\% sequence identity in the N terminus between the \textit{Drosophila} receptor and the mammalian receptors (Fig. 3). (iii) The N terminus of the \textit{Drosophila} receptor contains 9–10 Leu (Ile/Val)-rich repeats that occur in a similar number in the N termini of the mammalian receptors (Figs. 3 and 4). The number of those aliphatic residues in each repeating segment that can be aligned with the aliphatic residues of the other segments (which is nearly always 5; see Fig. 4A) and also the spacings between these aliphatic residues of the \textit{Drosophila} receptor N terminus are either the same or very similar to those found in the mammalian receptors (Fig. 4B; cf. Refs. 12 and 23). The
function of the Leu-rich repeats in the mammalian receptors has not been established yet. Recently, however, two research groups (12, 13) have developed two similar, three-dimensional models for the Leu-rich repeats in the N termini of various mammalian glycoprotein hormone receptors. These models are based on the crystal structure of the porcine ribonuclease-ribonuclease inhibitor complex, where the Leu-rich repeats of the inhibitor are arranged in a horseshoe-like structure and where these repeats function as the ribonuclease binding motifs (25, 26). In the models of Refs. 12 and 13, the Leu-rich repeats of the mammalian glycoprotein hormone receptor N terminus each have an inwardly directed β-sheet (at the concave surface) and an outwardly directed α-helix (at the convex surface of the horseshoe). The inwardly directed β-sheet structures are supposed to bind the ligand. The N terminus of the Drosophila receptor might have a similar configuration and ligand-binding capacity. Thus, the Drosophila receptor might not only be structurally, but also functionally, related to the mammalian receptors. (iv) There exists 50–53% sequence identity between the membrane-spanning domain of the Drosophila receptor and those of the mammalian glycoprotein hormone receptors. This sequence identity is very high if one takes into consideration that the sequence identity between the membrane-spanning regions of the rat LH/CG and TSH receptor is 70% and that the identity between the rat TSH and FSH receptor is 67% (10).

In addition to the structural homologies between the Drosophila and mammalian receptor proteins, the organization of their genes is also similar. Seven introns (introns 3–9; Fig. 1) in the DNA coding for the extracellular N terminus of the Drosophila receptor occur at exactly the same positions as in the genes coding for the mammalian receptors (indicated by the aliphatic residues marked by filled circles in Fig. 3). The triplets coding for the aliphatic residues at these seven intron-exon transitions are always interrupted after the second nucleotide (intron phase = 2; see Table I). The same intron phasing is found in the corresponding introns of the mammalian receptor genes (23, 27–29). This strongly suggests that the Drosophila receptor is evolutionarily related to the TSH, FSH, and LH/CG receptors from mammals.

The Drosophila receptor gene has also properties that are different from the mammalian receptor genes. The most striking difference is the existence of five introns in the region coding for the transmembrane area and one intron in the region coding for the intracellular C terminus (Figs. 1 and 2; Table I). In the genes coding for the mammalian glycoprotein hormone receptors, introns have only been found in the region coding for the large, extracellular N terminus (23, 27–29). For some of the other G protein-coupled receptor genes, however, it is known that introns do also occur in the regions coding for the seven-transmembrane domain and the intracellular C terminus, and these receptor genes have been found both in vertebrates and invertebrates, among them Drosophila (30, 31). It has always been assumed that the mammalian glycoprotein hormone receptors have originated from a recombination of a DNA sequence containing several exons coding for the repeating Leu-rich elements and an intronless sequence with one exon coding for a G protein-coupled receptor (23, 27–29). Our present findings now suggest that the gene for this ancestral G protein-coupled receptor had at least five introns (if one agrees with the theory that the number of introns in a certain gene does not increase but tends to decrease in the course of euaroyte evolution (32)). The intron-exon pattern in the gene region coding for the transmembrane and intracellular domains of the Drosophila receptor does not resemble that of any other known G protein-coupled receptor gene, indicating that the putative ancestral G protein-coupled receptor is not closely related to any of the known receptors that have been characterized at the genomic level so far. The recombination event creating the family of glycoprotein hormone receptors must have occurred early in evolution, since receptors similar to mammalian glycoprotein hormone receptors can already be found in sea anemones (14). In sea anemones, the genomic organization of the hormone receptor has only been partly investigated, and two introns have been found in the DNA region coding for the N terminus (at precisely the same positions as in the mammalian and Drosophila receptors; cf. Fig. 3). In light of our findings in Drosophila, it would be interesting to focus on the genomic organization of the region coding for the transmembrane domain of the sea anemone receptor to see whether this has indeed five or more introns.

When we started our research on the glycoprotein hormone receptor in Drosophila, it was our expectation to obtain a hormone receptor that was functionally related to the mammalian LH or FSH receptors and, thereby, to gain a better insight into the molecular mechanisms of insect reproduction. However, the developmental regulation of the novel Drosophila receptor (Fig. 8) clearly shows that receptor expression is not confined to sexually mature animals. Instead, high levels of mRNA are already produced in Drosophila embryos 8–16 h after oviposition, which is at a stage where the gonads have just started their embryonic development (33). This indicates that the Drosophila receptor must be involved in some developmental process. In this respect, it is interesting to know that mammalian LH and its receptor appear to play a role in Leydig cell proliferation and differentiation (34) and also that TSH stimulates the proliferation of mammalian thymocytes (35). In amphibians, a TSH-related protein hormone induces metamorphosis in tadpoles (36). These are all examples that glycoprotein hormone receptors can regulate development. On the other hand the possibility cannot be excluded that the Drosophila receptor is also involved in sexual reproduction, since adult male flies still have a high level of receptor mRNA (Fig. 8). In contrast to males, adult female flies have 6 times less receptor mRNA, which could point to a role of the receptor in spermagenesis or other male-specific reproductive processes. Future studies on the localization of the receptor mRNA and on the identification of the receptor ligand will help us to understand all of the functions of the novel insect receptor.

Acknowledgments—We thank Christina Faerch-Jensen for technical assistance, Astrid Juel Jensen and Marie Hansen for typing the manuscript, Dr. Leif Sondergaard (Department of Genetics, University of Copenhagen, Denmark) for support with setting up the Drosophila culture, and Drs. Eckart Gundelinger, Boupheng Phannavong, and Ulrich Thomas (Federal Institute for Neurobiology, Magdeburg, Germany) for hospitality and help with the localization of the receptor gene on salivary gland chromosomes.

REFERENCES

1. Libert, F., Lefort, A., Gérard, C., Parmentier, M., Perret, J., Ludygate, M., Dumont, J. E., and Vassart, G. (1989) Biochem. Biophys. Res. Commun. 165, 1250–1255
2. Losselt, H., Misrahi, M., Atger, M., Salesse, R., Vu Hai-Lau Thi, M. T., Jolivet, A., Guihoon-Mantel, A., Sar, S., Jallal, B., Garnier, J., and Milgrom, E. (1989) Science 245, 525–528
3. McFarland, K. C., Sprengel, R., Philips, H. S., Kohler, M., Rosenblit, N., Nikolics, K., Segaloff, D. L., and Seeburg, P. H. (1989) Science 245, 494–499
4. Parmentier, M., Libert, F., Maenhaut, C., Lefort, A., Gerard, C., Perret, J., van Sande, J., Dumont, J. E., and Vassart, G. (1989) Science 246, 1620–1622
5. Minegish, T., Nakamura, K., Takakura, Y., Miyamoto, K., Hasegawa, Y., Ihuki, Y., and Igarashi, M. (1990) Biochem. Biophys. Res. Commun. 172, 1049–1054
6. Minegish, T., Nakamura, K., Takakura, Y., Ihuki, Y., and Igarashi, M. (1991) Biochem. Biophys. Res. Commun. 175, 1125–1130
7. Misrahi, M., Losselt, H., Atger, M., Sar, S., Guihoon-Mantel, A., and Milgrom, E. (1990) Biochem. Biophys. Res. Commun. 166, 394–403
8. Sprengel, R., Braun, T., Nikolics, K., Segaloff, D. L., and Seeburg, P. H. (1990) Mol. Endocrinol. 4, 525–530
9. Kelton, C. A., Cheng, S. Y. Y., Nugent, N. P., Schweikhardt, R. L., Rosenthal, J. L., Overton, S. A., Wands, G. D., Kuzseja, J. B., Luchette, C. A., and...
Drosophila Glycoprotein Hormone Receptor

Chapel, S. C. (1992) *Mol. Cell. Endocrinol.* **89**, 141–151
10. Salesse, R., Remy, J. J., Levin, J. M., Jailal, B., and Garnier, J. (1991) *Biochimia* **73**, 109–120
11. Nagayama, Y., and Rapoport, B. (1992) *Mol. Endocrinol.* **6**, 145–156
12. Jiang, X., Dreano, M., Buckler, D. R., Cheng, S., Ythier, A., Wu, H., Hendrickson, W. A., and El Tayar, N. (1995) *Structure* **3**, 1341–1353
13. Kajava, A. V., Vassart, G., and Wodak, S. J. (1995) *Structure* **3**, 867–877
14. Nothacker, H.-P., and Grimmelikhuijzen, C. J. P. (1993) *Biochem. Biophys. Res. Commun.* **197**, 1062–1069
15. Tensen, C. P., van Kesteren, E. R., Planta, R. J., Cox, K. J. A., Burke, J. F., van Heerikhuizen, H., and Vreugdenhil, E. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4816–4820
16. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
18. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1988) *Science* **239**, 487–491
19. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
20. O’Connell, P. O., and Rosbash, M. (1984) *Nucleic Acids Res.* **12**, 5495–5513
21. Pardue, M. L. (1986) in *Drosophila: A Practical Approach* (Roberts, D. B., ed.), pp. 111–137, IRL Press, Oxford
22. Heckert, L. L., Daley, I. J., and Griswold, M. D. (1992) *Mol. Endocrinol.* **6**, 70–80
23. Mount, S. M. (1982) *Nucleic Acid Res.* **10**, 459–472
24. Koo, Y. B., Ji, I., Slaughter, R. G., and Ji, T. H. (1991) *Endocrinol.* **128**, 2297–2308
25. Tsai-Morris, C. H., Buczko, E., Wang, W., Xie, X-Z, and Dufau, M. L. (1991) *J. Biol. Chem.* **266**, 11355–11359
26. Gross, B., Misrahi, M., Sar, S., and Milgrom, E. (1991) *Biochem. Biophys. Res. Commun.* **177**, 679–687
27. Kremer, H., Kraaij, R., Toledo, S. P. A., Post, M., Fridman, J. B., Hayashida, C. Y., van Reen, M., Milgrom, E., Ropers, H.-H., Martinez Arias, A., and Brunner, H. G. (1995) *Nat. Genet.* **9**, 160–164
28. Vassart, G., and Dumont, J. E. (1992) *Endocr. Rev.* **13**, 596–611
29. Buckhinder, L., and Brown, D. D. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3820–3824
30. Kerrebrock, A. W., Moore, D. P., Wu, J. S., and Orr-Weaver, T. L. (1995) *Cell* **83**, 247–256