Second gene for gonadotropin-releasing hormone in humans

RICHARD B. WHITE*†, JONATHAN A. EISEN‡, THOMAS L. KASTEN*, AND RUSSELL D. FERNALD*

*Neuroscience Program and ‡Department of Biological Sciences, Stanford University, Stanford, CA 94305

Communicated by Donald Kennedy, Stanford University, Stanford, CA, November 11, 1997 (received for review July 15, 1997)

Gonadotropin-releasing hormone (GnRH) is ABSTRACT a decapeptide widely known for its role in regulating reproduction by serving as a signal from the hypothalamus to pituitary gonadotropes. In addition to hypothalamic GnRH (GnRH-I), a second GnRH form (pGln-His-Trp-Ser-His-Glv-Trp-Tyr-Pro-Gly; GnRH-II) with unknown function has been localized to the midbrain of many vertebrates. We show here that a gene encoding GnRH-II is expressed in humans and is located on chromosome 20p13, distinct from the GnRH-I gene that is on 8p21-p11.2. The GnRH-II genomic and mRNA structures parallel those of GnRH-I. However, in contrast to GnRH-I, GnRH-II is expressed at significantly higher levels outside the brain (up to 30×), particularly in the kidney, bone marrow, and prostate. The widespread expression of GnRH-II suggests it may have multiple functions. Molecular phylogenetic analysis shows that this second gene is likely the result of a duplication before the appearance of vertebrates, and predicts the existence of a third GnRH form in humans and other vertebrates.

Gonadotropin-releasing hormone (GnRH) is a decapeptide widely known for its role in regulating reproduction. Release of GnRH from the hypothalamus controls the production of pituitary gonadotropins responsible for gonadal development and growth in all vertebrates. This function for GnRH has been highly conserved during 500 million years of vertebrate evolution despite the fact that its amino acid sequence varies by 50% (1).

In addition to the hypothalamic GnRH of variable sequence, many vertebrate species have been shown to express a second, invariant GnRH form (pGln-His-Trp-Ser-His-Gly-Trp-Tyr-Pro-Gly; GnRH-II) (2). By using antibody staining, this form of GnRH has been found in the midbrain in all species where its location has been described (reviewed in ref. 1). Furthermore, nucleic acid probes have been used to identify GnRH-II expression in the midbrain of several fish species and one mammal (1).

Recently, a cDNA encoding this second form of GnRH was found in a placental mammal, the tree shrew *Tupaia glis* (1), thus leading us to search for it in humans. Here we describe the cloning of a cDNA encoding a second form of GnRH in humans and the subsequent isolation and sequencing of the complete human GnRH-II gene, the first description of a nonhypothalamic GnRH gene form in any species. In addition, the structure and chromosomal location of the new GnRH-II gene is compared with that of the previously described form in humans. Finally, we have constructed a molecular phylogeny of GnRH evolution, incorporating new sequence data for GnRH-II cDNAs from three placental mammals: human (this paper), tree shrew (1), and musk shrew (*Suncus murinus*; R.B.W., T.L.K., S. White, and R.D.F., unpublished data).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

@ 1998 by The National Academy of Sciences 0027-8424/98/95305-5\$2.00/0 PNAS is available online at http://www.pnas.org.

MATERIALS AND METHODS

Library Screen. A 270-nt partial cDNA for the putative human GnRH-II was cloned from human thalamus poly(A) RNA (CLONTECH) by using reverse transcription–PCR (RT-PCR) and 3'-RACE (rapid amplification of cDNA ends) as described (1). Oligomers flanking putative intron B were used to screen a human genomic P1 artificial chromosome (PAC) library (Genome Systems, St. Louis) by using PCR, and a single PAC clone of ≈100 kb containing the entire gene for GnRH-II was identified.

Sequence Analysis. Two overlapping PAC subclones spanning 4,498 bp, from 1.3 kb upstream to 1.1 kb downstream of the GnRH-II gene, were sequenced on both strands. Exons were predicted by using GRAIL (Oak Ridge National Laboratory Informatics Group), FGENEH, HSPL (V. Solovyev, Baylor College of Medicine), and NNSSP (M. Reese, Lawrence Berkeley National Laboratory), and by comparison with our cDNA information. Promoter predictions were made with GRAIL and NNPP (M. Reese, Lawrence Berkeley National Laboratory).

Fluorescence in Situ Hybridization Mapping. The PAC clone was labeled with digoxygenin dUTP by nick translation. Labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes in a solution containing 50% formamide, 10% dextran sulfate and 2× standard saline citrate (SSC; 1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7). A biotin-labeled probe specific for the centromere of chromosome 20 (D20Z1) was cohybridized with the PAC probe, and 71 of 80 metaphase cells analyzed exhibited specific labeling.

RNA Blot Hybridizations. A dot blot of human RNA (MasterBlot, CLONTECH) was used to quantify expression levels of GnRH-II. The blot consisted of poly(A) RNAs from 50 human tissues, with loadings of 80-448 ng per dot, normalized for eight housekeeping genes (7, 8), along with appropriate positive and negative controls. Samples on the blot were taken from whole brain, amygdala, caudate nucleus, cerebellum, cerebral cortex, frontal lobe, hippocampus, medulla oblongata, occipital lobe, putamen, substantia nigra, temporal lobe, thalamus, subthalamic nucleus, spinal cord, heart, aorta, skeletal muscle, colon, bladder, uterus, prostate, stomach, testis, ovary, pancreas, pituitary gland, adrenal gland, thyroid gland, salivary gland, mammary gland, kidney, liver, small intestine, spleen, thymus, peripheral leukocyte, lymph node, bone marrow, tonsil, lung, trachea, placenta, fetal brain, fetal heart, fetal kidney, fetal liver, fetal spleen, fetal thymus, fetal lung, yeast total RNA, yeast tRNA, Escherichia coli rRNA, E. coli DNA, poly r(A), human cot1 DNA, and human DNA. Hybridization was overnight at 65°C in ExpressHyb (CLONTECH) using a GnRH-II cDNA probe and final

Abbreviations: GnRH, gonadotropin-releasing hormone; RT-PCR, reverse transcription-PCR; PAC, P1 artificial chromosome; GAP, GnRH-associated peptide.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF036329 and AF036330).

[†]To whom reprint requests should be addressed at: Neuroscience Program, Building 420, Stanford University, Stanford, CA 94305-2130. e-mail: rwhite@psych.stanford.edu.

washing was at 55°C in 0.1× SSC/0.5% SDS. The blot was exposed to a phosphor screen and the hybridization signals were quantified by using IPLabGel (Signal Analytics, Vienna, VA). The blot was then stripped and rehybridized (under the same conditions) with a 35-nt oligomer complementary to the GnRH-associated protein (GAP) sequence in exon 3 (5'-CC-CAGGGCATGCTGTCGTCCAGGGGAGCCAGGGCA-3'). Hybridization with GnRH-II probes of a second dot blot, consisting of RNA from different individuals, yielded similar results. Hybridization with cDNA and oligomer probes complementary to GnRH-I revealed no significant regions of expression outside the brain. Tissue sources for the poly(A) RNAs with high levels of GnRH-II expression were: (*i*) kidney, pooled from one male, age 49, and one female, age 63; (ii) prostate, pooled from 35 males ages 21-70; and (iii) bone marrow, pooled from 120 males and females ages 16-70.

A Northern blot of human brain tissues (2 μ g poly(A) mRNA per lane) (CLONTECH) was hybridized with either a GnRH-I-specific or GnRH-II-specific cDNA probe. Hybridization was at 65°C for 1.5 h in ExpressHyb and final washing was at 50°C in 0.1× SSC/0.1% SDS. Rehybridization with a probe for β -actin was used to verify RNA integrity and equivalent loading (not shown).

RT-PCR. Fetal brain, adult thalamus, and adult kidney poly(A) RNA (CLONTECH) were transcribed into cDNA by using an oligo-dT₁₈ primer and Superscript II RT (GIBCO/Life Technologies, Grand Island, NY), and then amplified by PCR for 30 cycles by using oligomers complementary to exon 1 (5'-CTGCAGCTGCCTGAAGGAG-3') and exon 4 (5'-GG-GCGGGGGGGGGCTCTCG-3') of GnRH-II. Confirmation as human GnRH-II RNA products was by probing a Southern blot with the 35-mer described above, and by sequencing. Kidney poly(A) RNA source was a female, age 65.

Phylogenetic Analyses. The sequences of GnRH-preprohormones were aligned to each other by using CLUSTAL W (9) with default settings. A central core domain including part of the signal sequence, all of GnRH, and part of the GAP sequence was used for the analyses. Partial sequences that did not include this entire domain were excluded. For distancebased methods, pairwise distances between each protein were calculated by using PROTDIST in PHYLIP (10). These distances were used as input for the neighbor-joining tree reconstruction algorithm (11) in PHYLIP. Maximum parsimony trees were generated by using PROTPARS in PHYLIP. Maximum likelihood trees were generated by using PUZZLE (12). Bootstrap values, indicating the number of times a particular set of sequences group together when trees are generated from resampled alignments, were calculated by using the method of Felsenstein in PHYLIP (13). Full species names and GenBank accession numbers for the different GnRH types (I, II, and III) are as follows: Human (Homo sapiens) I: X01059, II: tree shrew

(Tupaia glis belangeri) I: U63326, II: U63327; musk shrew (Suncus murinus) R.B.W., T.L.K., S. White, and R.D.F., unpublished data; pig (Sus scrofa) I: L32864; mouse (Mus musculus) I: M14872; rat (Rattus norvegicus) I: M15527; chicken (Gallus gallus) I: X69491; frog (Xenopus laevis) I: L28040, gilthead sea bream (Sparus aurata) I: U30320, II: U30325, III: U30311; cichlid (Haplochromis burtoni) I: U31865, II: L27435, III: S63657; red sea bream (Chrysophrys major) I: D86582, III: D26108; African catfish II: X78047, IIB:X78048; goldfish (Carassius auratus) II: U30386, III: U30301; plainfin midshipman (Porichthys notatus) III: U41669; Salmonids A III [Chinook salmon (Oncorhynchus tschawytscha) X79711; rainbow trout (Oncorhynchus mykiss) X79710; Atlantic salmon (Salmo salar) X79709; sockeye salmon (Oncorhynchus nerka) IIIa: D31869; brown trout (Salmo trutta) X79713; brook trout (Salvelinus fontinalis) X79712]; Salmonids B [cherry salmon (Oncorhynchus masou) D10946; sockeye salmon IIIb: D31868]. Supporting data sets: http://wwwleland.stanford.edu/~jeisen/GnRH/GnRH.html.

RESULTS AND DISCUSSION

To discover whether humans have a second GnRH form, we used RT-PCR on human thalamic mRNA and obtained a partial cDNA encoding a putative second GnRH. This cDNA was used to design oligonucleotide primers to screen a human genomic PAC library with PCR. A single clone was isolated and subsequent sequence analysis confirmed that it contained the complete gene encoding GnRH-II. We then used fluorescence in situ hybridization to map GnRH-II unambiguously to the p terminus of chromosome 20, corresponding to band 20p13, different from GnRH-I, which is located at 8p21-p11.2 (14). Additionally, the human GnRH-II gene is remarkably short (2.1 kb) compared with GnRH-I (5.1 kb) (15) (Fig. 1), primarily because introns B and C are much larger in GnRH-I. However, both the new GnRH-II gene and the human GnRH-I gene (16) have four exons, as do GnRH genes in other species (2), and the predicted preprohormone encoded by human GnRH-II is organized identically to that of other GnRHs: all have a signal sequence, followed by a GnRH decapeptide, a conserved proteolytic site (Gly-Lys-Arg), and a GAP. Corresponding preprohormone elements of the two human GnRHs are quite similar in length, with the striking exception that GAP is 50% longer in GnRH-II than in GnRH-I (84 vs. 56 amino acids). A similar disparity in GAP was recently reported for tree shrew (76 vs. 56 amino acids; ref. 1), suggesting that a relatively larger GAP may be common among mammalian GnRH-II forms.

Surprisingly, human GnRH-II mRNA expression is highest outside the brain. Of 50 tissues tested, expression levels in the kidney were \approx 30-fold higher than in any brain region, and

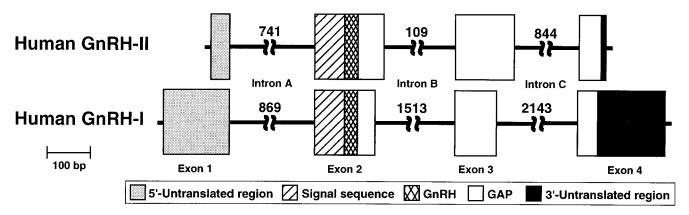


FIG. 1. Organization of the human GnRH-II gene compared with that of human GnRH-I (16). Note that only exonic regions are drawn to scale. Intron sizes in bp are indicated. GnRH-II exon 1 size is based on predicted location of the most proximate promoter.

expression in bone marrow and prostate was ≈4-fold greater than in the brain. GnRH-I expression was not observed at a high level outside the brain. Within the brain, GnRH-II expression is most obvious in the caudate nucleus but can also be seen in the hippocampus and amygdala (Fig. 2a). Hybridization of the same blot with a probe specific for human GnRH-I showed a wholly different pattern (Fig. 2b), confirming that cross-hybridization cannot account for these results. Further confirmation was obtained by using RT-PCR to amplify GnRH-II mRNA in three tissues (thalamus, fetal brain, and kidney). Each yielded a cDNA fragment of the predicted size, and Southern blot analysis established these as human GnRH-II RNA products (Fig. 3). The blot revealed the existence of a GnRH-II cDNA splice variant extended at the 5'-end of exon 3, such that the predicted length of GAP is increased from 77 to 84 amino acids. This longer GnRH-II transcript was found in both fetal brain and adult thalamus, but not in adult kidney, implying tissue-specific processing of GnRH-II transcripts.

Medical Sciences: White et al.

Our discovery of a GnRH-II gene in humans, in combination with recent GnRH-II cDNA data for tree shrew (1) and musk shrew (R.B.W., T.L.K., S. White, and R.D.F., unpublished data), led us to examine the evolution of GnRH peptides. Molecular phylogenetic reconstruction was used to infer the evolutionary relationship among GnRH forms (Fig. 4). The methods of maximum parsimony, maximum likelihood, and neighbor-joining were used to test the robustness of our analysis, and each method showed essentially the same branching patterns. The phylogenetic tree shows the existence of three evolutionarily distinct GnRH groups: "releasing" forms localized to the hypothalamus (GnRH-I), forms previously localized solely to midbrain nuclei (GnRH-II), and forms localized to the telencephalon in several fish species (GnRH-III).

Several lines of evidence support the conclusion that the three GnRH groups identified by this phylogenetic comparison comprise evolutionarily distinct forms. First, the groupings correspond to GnRH forms with distinct expression patterns

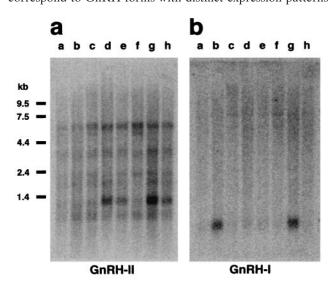


FIG. 2. (A) Expression of human GnRH-II in the brain. A Northern blot (CLONTECH) of poly(A) mRNA from selected brain regions was hybridized with a GnRH-II cDNA probe. Tissues: a, thalamus; b, subthalamic nucleus; c, substantia nigra; d, whole brain; e, hippocampus; f, corpus callosum; g, caudate nucleus; h, amygdala. mRNA sizes are indicated. A 1.3-kb transcript was most abundant, although a 6.5-kb transcript was consistently observed under a variety of hybridization conditions. No transcript was seen at a size (\approx 0.6 kb) predicted by use of the most proximate promoter. (B) Expression of human GnRH-I in the brain. GnRH-I cDNA was used to re-probe the same blot as in A. Only a single, \approx 0.7-kb transcript was observed, as expected (27).

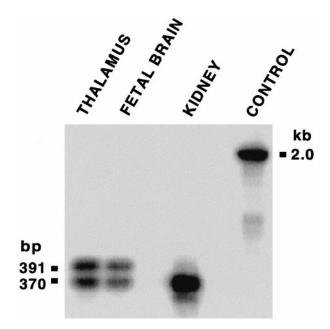


Fig. 3. Confirmation of GnRH-II mRNA in thalamus, fetal brain, and kidney by RT-PCR and demonstration of a truncated transcript in brain. Reaction products were prepared as a Southern blot and hybridized with an oligonucleotide specific to a sequence between the amplification primers. Control amplification of human genomic DNA yielded a product of the expected size.

and thus presumed biological roles. This suggests the groupings can be used to predict the activities of yet uncharacterized forms. Second, although only the neighbor-joining tree is shown, the same groupings were found by using other distance-based methods as well as maximum parsimony and maximum likelihood (not shown). Third, high bootstrap values were generated for the neighbor-joining and parsimony trees, strongly supporting identification of these groups. Finally, the branch lengths to each group are relatively long, suggesting an ancient divergence. Furthermore, the branching order within each GnRH group matches the evolutionary branching order of the species (see in particular GnRH-I, for which sequences were available from diverse species).

Although these phylogenetic relationships cannot identify the selective forces that generated multiple GnRH forms, the analysis can suggest when distinct forms arose. The structure of this phylogenetic tree strongly suggests that multiple forms exist in many different species because of ancient duplications of a gene encoding GnRH. Moreover, it seems likely that different forms of GnRH diverged from each other before the divergence of species represented in the tree. Because the tree is unrooted, the exact order of the duplications in unclear. However, GnRH-I and GnRH-II each include representatives from fish and mammals, so the separation of these two groups must have occurred before the separation of mammalian and fish ancestors. The origin of GnRH-III is less clear, because there are only representatives from fish in this group, but two alternative hypotheses are suggested by our data. First, it is possible that the gene duplication leading to GnRH-III occurred only recently, within the fish lineage, and thus is restricted to this taxon. Alternatively, GnRH-III may have resulted from an ancient duplication, and has either been lost in higher vertebrates during evolution or has not yet been discovered. We favor the latter hypothesis because if the duplication were recent, the GnRH-III forms would be expected to cluster with one of the other GnRH forms in fish. Perhaps another form of GnRH remains to be found in

Among the GnRH decepeptides, only GnRH-II is identical in all species in which it has been found. This suggests that the

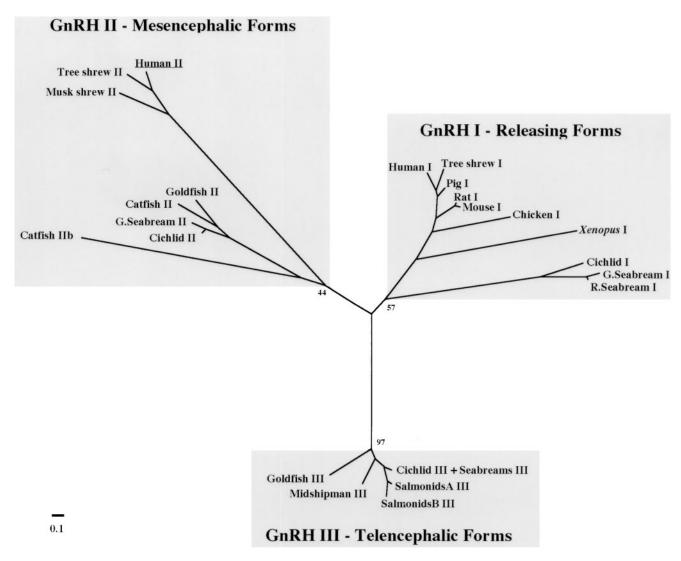


FIG. 4. Unrooted neighbor-joining phylogenetic tree of GnRH peptides. The tree was generated from an alignment of GnRH preprohormones by using the neighbor-joining algorithm (11) and the Hall distance calculation correction procedure in PHYLIP (10). Evolutionary distances are represented only by the length of the branches and not by branch angles. Bootstrap values, indicating the number of times a particular set of sequences group together when trees are generated from resampled alignments, are indicated for some important nodes on the tree. An unrooted tree is shown because of the lack of an obvious outgroup. Some GnRH-III peptide sequences from different species were identical and are grouped together. The scale bar corresponds to estimated evolutionary distance units as calculated by PROTDIST in PHYLIP. For accession numbers and species names, see *Materials and Methods*.

peptide has been subject to extremely stringent selective pressures. Such selection might have arisen if GnRH-II serves different functions at different loci in the body, as supported by several examples. First, GnRH-II has been shown to be present in the sympathetic ganglia of amphibia (17) where it has been shown that GnRH can act as a neuromodulator in spinal cord ganglion neurons in amphibia (3). In fish, neurons shown previously to control sperm duct and oviduct contractility receive input from cells that show GnRH-like immunoreactivity (5). Moreover, in newts, immunoreactivity to GnRH-II shifts from midbrain cell bodies to terminal regions following the initiation of courtship (4). Examples of possible GnRH-II expression outside the midbrain cells include transient GnRH immunoreactivity within mast cells in the habenula of ring doves following courtship (18, 19). Similarly, mast cells in musk shrew pup brains (6) are immunoreactive for GnRH by using an antibody shown to be specific to GnRH-II. These data have led to the suggestion that mast cells represent an alternative delivery system for GnRH (18). Indeed, several types of differentiated lymphocytes including spleenocytes, thymocytes, peripheral T- and B-lymphocytes, and mast cells

(20) have been demonstrated to produce GnRH or a GnRH-like peptide. Because kidney, bone marrow, and prostate contain significant numbers of mast cells (21–23), perhaps these cells account for the high GnRH-II expression there. Furthermore, although it is not known whether different forms of GnRH might have different receptor types, GnRH receptors have been found throughout the body, including the kidney (24) and prostate (25). Because previous studies have shown that GnRH receptors bind GnRH-II peptide up to 100 times more effectively than GnRH-I (26), GnRH-II could act through these receptors outside the brain. Perhaps GnRH-II has multiple functions as suggested by its presence in diverse loci.

Has GnRH-II remained unchanged because of selective pressure from multiple functions? Perhaps GnRH-II originated in the immune system and only later acquired a neuromodulatory function in the brain. Additional characterization of GnRH-II should help clarify whether it serves distinct functions in the body and what selective forces are responsible for its remarkable conservation over evolutionary time.

We thank D. Anderson, E. Gestrin, H. Hofmann, K. Hoke, R. Robison, and S. White for comments on the manuscript, L. Clark and

K. Wilhelmsen for use of the PAC library, U. Demarco and M. Benson for technical assistance, E. Rissman for musk shrew tissue, J. Adelman for GnRH-I cDNA, and J. Felsenstein for providing PHYLIP free of charge. R.B.W., T.L.K., and R.D.F. were supported by National Institutes of Health Grant NS 34950 and J.A.E. was supported by National Cancer Institute Grant CA44349.

- Kasten, T., White, S., Norton, T., Bond, C., Adelman, J. & Fernald, R. (1996) Gen. Comp. Endocrinol. 104, 7-19.
- Sherwood, N., Lovejoy, D. & Coe, I. (1993) Endocr. Rev. 14, 241-254.
- Jan, Y. N., Jan, L. Y. & Brownfield, M. S. (1980) Nature (London) **288**, 380-382.
- Muske, L. E., King, J. A., O'Connell, B. G., Moore, F. L. & Millar, R. P. (1995) Soc. Neurosci. Abstr. 21, 100.
- Miller, K. E. & Kriebel, R. M. (1986) Gen. Comp. Endocrinol. 64, 396 - 400.
- Rissman, E. F., Alones, V. E., Craig-Veit, C. B. & Millam, J. (1995) J. Comp. Neurol. 357, 524-531.
- Spanakis, E. (1993) Nucleic Acids Res. 21, 3809-3819.
- Spanakis, E. & Brouty-Boye, D. (1994) Nucleic Acids Res. 22, 799-806.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680.
- Felsenstein, J. (1989) Cladistics 5, 164-166.
- Saitou, N. & Nei, M. (1987) Mol. Biol. Evol. 4, 406-425. 11.
- Strimmer, K. & von Haeseler, A. (1996) Mol. Biol. Evol. 13, 12.
- Felsenstein, J. (1985) Evolution 39, 783-791. 13.

- Yang-Feng, T. L., Seeburg, P. H. & Francke, U. (1986) Somatic Cell Mol. Genet. 12, 95-100.
- Hayflick, J. S., Adelman, J. P. & Seeburg, P. H. (1989) Nucleic Acids Res. 17, 6403-6404.
- Adelman, J., Mason, A., Hayflick, J. & Seeburg, P. (1986) Proc. Natl. Acad. Sci. USA 83, 179-183.
- Troskie, B., King, J. A., Millar, R. P., Peng, Y., Kim, J., Figueras, H. & Illing, N. (1997) Neuroendocrinology 65, 396-402.
- Silver, R., Ramos, C. L. & Silverman, A. J. (1992) J. Neuroendocrinol. 4, 207-210.
- Silverman, A. J., Millar, R. P., King, J. A., Zhuang, X. & Silver, R. (1994) Proc. Natl. Acad. Sci. USA 91, 3694-3699
- Marchetti, B., Gallo, F., Farinella, Z., Romeo, C. & Morale, M. C. (1996) Ann. N.Y. Acad. Sci. 784, 209-236.
- Assem, E. S. K., Abdullah, N. A. & Cowie, A. G. A. (1987) Int. Arch. Allergy Appl. Immunol. 84, 212-216.
- Orfao, A., Escribano, L., Villarrubia, J., Velasco, J. L., Cervero, C., Ciudad, J. & Navarro, J. L. (1996) Am. J. Pathol. 149, 1493-1499.
- Fritz, F. J. & Pabst, R. (1989) Int. Arch. Allergy Appl. Immunol. 88, 360-362
- Kakar, S. & Jennes, L. (1995) *Cancer Lett.* **98,** 57–62. Fekete, M., Redding, T. W., Comraru-Schally, A. M., Pontes, J. E., Conneley, R. W., Srkalovic, G. & Schally, A. V. (1989) Prostate 14, 191-208.
- 26. King, J. A. & Millar, R. P. (1991) in Vertebrate Endocrinology, Vol. 4: Fundamentals and Biomedical Implications, Part B, eds. Pang, P. & Schreibman, M. (Academic, San Diego), pp. 1-32.
- Seeburg, P. H. & Adelman, J. P. (1984) Nature (London) 311, 666 - 668.