Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 3043–3046, May 1987
Neurobiology

Cloning and sequence analysis of cDNAs for neurohypophysial hormones vasotocin and mesotocin for the hypothalamus of toad, *Bufo japonicus*

(recombinant DNA/amphibia/neurosecretory system/molecular evolution/neuropeptides)

HIROSHI NOJIRI*†, ISAO ISHIDA*, ETSUKO MIYASHITA‡, MORIYUKI SATO‡, AKIHISA URANO§¶, AND TAKEO DEGUCHI*

*Department of Molecular Neurobiology, Tokyo Metropolitan Institute for Neurosciences, 2-6 Musashidai, Fuchu-city, Tokyo 183; ‡Tokyo Research Laboratories, Kyowa Hakko Kogyo Machida, Tokyo 194; and §Department of Regulation Biology, Faculty of Science, Saitama University, Urawa, Saitama 338, Japan

Communicated by Howard A. Bern, December 29, 1986 (received for review November 20, 1986)

ABSTRACT **The primary structures of the precursors of neurohypophysial hormones vasotocin (VT) and mesotocin (MT) in the hypothalamus of the toad *Bufo japonicus* were determined by analyzing the nucleotide sequences of the cloned cDNAs encoding them. The MT precursor consists of 125 amino acid residues containing a signal peptide followed directly by MT, which in turn is connected to the MT neurophysin by Gly-Lys-Arg, a processing and carboxyl-terminal amidation signal. In contrast, the VT precursor includes a glycoprotein of 36 amino acids following the VT neurophysin. Except for glycoprotein, the structures of MT and VT precursors are quite similar. RNA transfer blotting analysis showed that both MT and VT mRNAs are present in the brain but not in the liver, ovaries, and testes of the toad. The sequences and the structural organizations of the MT and VT precursors are highly homologous to those of their mammalian counterparts, oxytocin and arginine vasopressin precursors, respectively. This fact suggests that, in the evolutionary pathway of neurohypophysial hormones, VT is the ancestor molecule of vasopressin, while MT is that of oxytocin.**

Vasotocin (VT) and mesotocin (MT) are neurohypophysial peptides produced by magnocellular preoptic neurons in the hypothalamus in amphibians, reptiles, and birds. They have important physiological roles in the regulation of water balance in anuran amphibians (1). VT is probably an antidiuretic hormone, while MT may be a diuretic one. Recent neuroethological studies have suggested a further physiological significance of VT in anuran reproductive behavior (2).

Ten distinct nonapeptide principles have been characterized in a wide variety of vertebrates (3). The molecular evolution of neurohypophysial hormones has attracted many investigators, and many schemes have been proposed for an evolutionary pathway based on their amino acid structures and phyletic distributions (4). The structures of precursors for neurohypophysial hormones in lower vertebrates, however, have not yet been characterized.

Recently, the primary structures of the precursors of mammalian neurohypophysial hormones, arginine vasopressin ([Arg^8]VP) and oxytocin (OT), were determined by sequence analysis of cDNAs from cows (5, 6). Thereafter, the gene organizations for [Arg^8]VP and OT were elucidated in cows (7), rats (8), and humans (9). According to these studies, the [Arg^8]VP precursor consists of ternary segments for [Arg^8]VP, [Arg^8]VP-neurophysin, a carrier protein specific to [Arg^8]VP, and a glycoprotein named copeptin. The OT precursor consists of binary segments for OT and OT-specific neurophysin lacking a glycoprotein domain. Since VT and MT are the most widely distributed neurohypophysial principles in lower vertebrates, elucidation of the structures of their precursors would provide valuable information as to the evolutionary relationship of these peptides.

We report here the isolation and characterization of the cDNAs encoding the mRNAs for VT and MT precursors of the toad, *Bufo japonicus*. The elucidated nucleotide sequences show considerable homology in both organizations and sequences among the four precursors of neurohypophysial hormones (VT, MT, [Arg^8]VP, and OT).

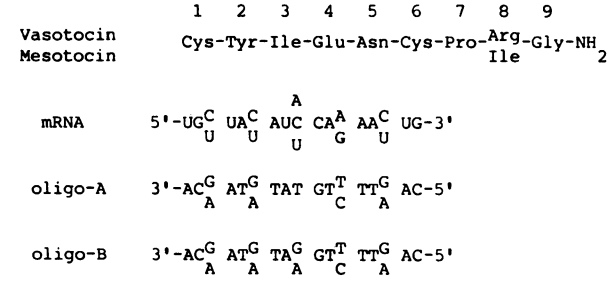

FIG. 1. Synthetic oligodeoxynucleotides for the screening of MT and VT coding sequences. The amino acid sequences and the predicted nucleotide sequences of mRNAs for the common region of MT and VT are shown. Based on these nucleotide sequences, oligo-A, which is a pool of 16 oligonucleotides, and oligo-B, a pool of 32 oligonucleotides, were synthesized by use of the phosphoramidite method (14) as probes for the screening.

MATERIALS AND METHODS

Procedures for constructing a cDNA library from the toad diencephalons, screening of the library, analyzing the nucleotide sequences, and RNA transfer-blot analysis followed those of Ishida *et al.* (10). In brief, the diencephalons (15 g) were collected from 300 toads captured in the breeding season in March 1985, and a total RNA (5.1 mg) was extracted from them in 4 M guanidinium thiocyanate buffer (11). Poly(A)$^+$ RNA (115 μg) was separated by oligo(dT)-cellulose column chromatography (12). A cDNA library was constructed by the method of Okayama and Berg (13) with 10 μg of poly(A)$^+$ RNA and 3.2 μg of vector/primer DNA. A total of 20,000 transformants was screened by colony hybridizations with two ^{32}P-labeled synthetic oligonucleotide mixtures designated "oligo-A" and "oligo-B" (Fig. 1), which are complementary to all putative mRNA sequences predicted

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.

Abbreviations: [Arg^8]VP, arginine vasopressin; MT, mesotocin; OT, oxytocin; VT, vasotocin.

†On leave from Saitama University. Present address: Tochigi First Lab, Kaou Co., Akabane, Ichikai-machi, Haga-gun, Tochigi 321-34, Japan.
¶To whom reprint requests should be addressed.

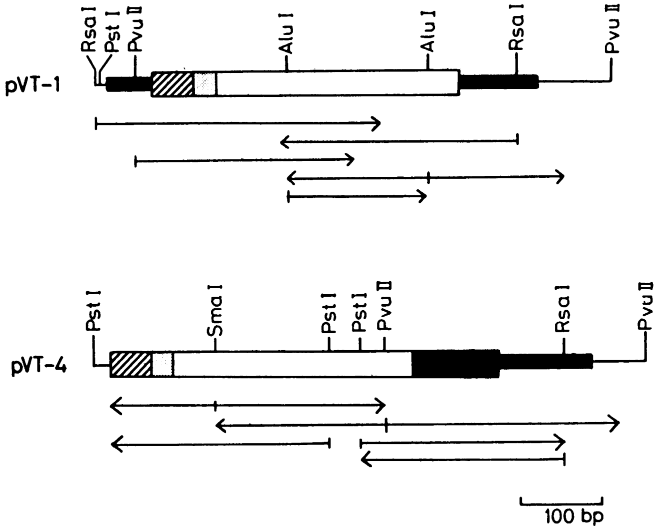

FIG. 2. Restriction maps and sequencing strategies of the toad MT (pVT-1) and VT (pVT-4) cDNA clones. From left to right, each diagram shows a schematic representation of the 5′ untranslated region (thick line), the signal peptide (hatched box), MT or VT (stippled box), the MT- or VT-specific neurophysin (open box), the glycoprotein region (heavily stippled box in pVT-4), and the 3′ untranslated region (thick lines). Horizontal arrows indicate the direction and extent of sequence determination.

from the amino acids sequence common to VT and MT. The probes were 5′-end-labeled with [γ-^{32}P]ATP to a specific activity of 2–3 × 10^7 cpm/μg. Plasmids from the positive clones were analyzed by several restriction enzymes. The inserts from the positive clones were subcloned into pUC18 vector, and the nucleotide sequences were determined by the Sanger dideoxy sequencing method (15).

Poly(A)$^+$ RNAs from the brains, testes, ovaries, and livers of the toads were electrophoresed in a 1.4% agarose/formaldehyde gel, and transferred to a nitrocellulose filter (16). After prehybridization, the filter was hybridized with a ^{32}P-labeled nick-translated *Pvu* II–*Rsa* I fragment of pVT-1 (MT cDNA clone) or a *Ava* I–*Ava* I fragment of pVT-4 (VT cDNA clone) and was subjected to autoradiography.

RESULTS

Cloning of cDNAs Encoding VT and MT Precursors. By screening of about 20,000 transformants with the 5′-^{32}P-labeled synthetic oligonucleotide probes, 2 positive colonies were obtained by oligo-A, and 4 by oligo-B. The former 2 colonies had inserts of approximately 600 base pairs. Digestion of these inserts with *Pvu* II and *Rsa* I yielded fragments of the same size, suggesting that they may be identical. The analysis of the nucleotide sequence of 1 clone (pVT-1) showed that it contained a MT-specific sequence. Meanwhile, a restriction analysis of the clones obtained by oligo-B showed a presence of two separate groups. One clone (pVT-4) included a VT-specific sequence, while 3 clones in another group did not contain any sequence specific to VT or MT.

Structural Analysis of MT and VT Precursors. Inserts of pVT-1 and pVT-4 were subjected to treatments with restriction endonucleases, and the nucleotide sequences were analyzed by the strategies illustrated in Fig. 2. The nucleotide sequences of mRNAs and the deduced amino acid sequences of the precursors of toad MT and VT are shown in Figs. 3 and 4.

MT Precursor. The pVT-1 insert is composed of 525 nucleotides encoding MT at nucleotides 1–27. The initiation site of translation may be assigned to the methionine codon AUG at positions −54 to −52, but not at positions −63 to −61, because CCGCCAUG at positions −59 to −52 is almost the same as the initiation site of the other eukaryotic genes [consensus sequence: CCRCCAUG (where R = purine)] (17). Comparison of the structure of the MT precursor with that of the OT precursor further supports it. The succeeding 18, mainly hydrophobic, amino acid residues probably represent the signal peptide. A tripeptide (Gly-Lys-Arg) that follows MT may serve as the signal for carboxyl-terminal amidation and proteolytic processing (18). The remainder of the MT precursor consists of a peptide of 94 amino acids that is cysteine-rich and highly homologous to the mammalian OT- and [Arg^8]VP-specific neurophysins (5, 6, 18). A termination codon (UAG) follows directly the sequence encoding the neurophysin-like peptide and is followed by a 3′ untranslated region containing an AAUAAA sequence (396–401) for its polyadenylylation signal. Thus, the precursor of toad MT is composed of 125 amino acid residues and has a calculated molecular weight of 13,516.

```
                                                                 -100
                                                     AGAAGAACUCCAAGGAACACUCAUCUUCGCA
                       -50                                                        -1
GACCAGCUGGAGUCAUGACCGCC AUG UCU UAC ACU GCA CUU GCC GUG ACC UUU UUU GGC UGG UUA GCC UUG UCU UCA GCC
                        Met Ser Tyr Thr Ala Leu Ala Val Thr Phe Phe Gly Trp Leu Ala Leu Ser Ser Ala
                        -19                                 -10                                 -1
1                                                                50
UGU UAU AUA CAA AAC UGC CCC AUA GGA GGA AAG AGA UCG GUG AUA GAC UUC AUG GAC GUA AGA AAG UGU AUC CCC
Cys Tyr Ile Gln Asn Cys Pro Ile Gly Gly Lys Arg Ser Val Ile Asp Phe Met Asp Val Arg Lys Cys Ile Pro
 1          MESOTOCIN               10           NEUROPHYSIN         20
                                   100                                                              150
UGU GGA CCA AGG AAC AAA GGU CAC UGC UUU GGG CCC AAC AUC UGC UGU GGA GAG GAG CUG GGU UGU UAC UUU GGC
Cys Gly Pro Arg Asn Lys Gly His Cys Phe Gly Pro Asn Ile Cys Cys Gly Glu Glu Leu Gly Cys Tyr Phe Gly
                30                                  40                                          50
                                                                 200
ACC ACG GAG ACU CUG AGA UGC CAA GAA GAA AAU UUC CUC CCU UCA CCG UGU GAA UCU GGA AGG AAA CCA UGU GGG
Thr Thr Glu Thr Leu Arg Cys Gln Glu Glu Asn Phe Leu Pro Ser Pro Cys Glu Ser Gly Arg Lys Pro Cys Gly
                                    60                                      70
                                   250                                                              300
AAU AAU GGA GGA AAC UGC GCC CGC UCU GGC AUC UGC UGC AAC CAU GAG AGC UGC ACG AUG GAC CCG GCC UGC GAA
Asn Asn Gly Gly Asn Cys Ala Arg Ser Gly Ile Cys Cys Asn His Glu Ser Cys Thr Met Asp Pro Ala Cys Glu
                80                                              90                              100
                                                            350
CAG GAC UCG GUA UUU UCC UAG UCCCUAGACCUUCUCCAAGGCAGAACCACCACACGAAGCGGGACCACCGCUGUGUAUAGAUGUACUGUACAG
Gln Asp Ser Val Phe Ser ***
                     106
     400
UGAAUAAAGACCUCAAACAUG-poly(A) 3'
```

FIG. 3. Primary structure of the mRNA from the toad brain encoding the MT precursor and the deduced amino acid sequence. Nucleotide residues are numbered in the direction from 5′ to 3′, beginning with the first residue in the coding region for MT. The nucleotides upstream to the residue 1 are indicated by negative numbers. The amino acid residues are numbered with the first residue (Cys) of MT as 1. The amino acids constituting the putative signal peptide are indicated by negative numbers. The AAUAAA sequence in the 3′ untranslated region is underlined.

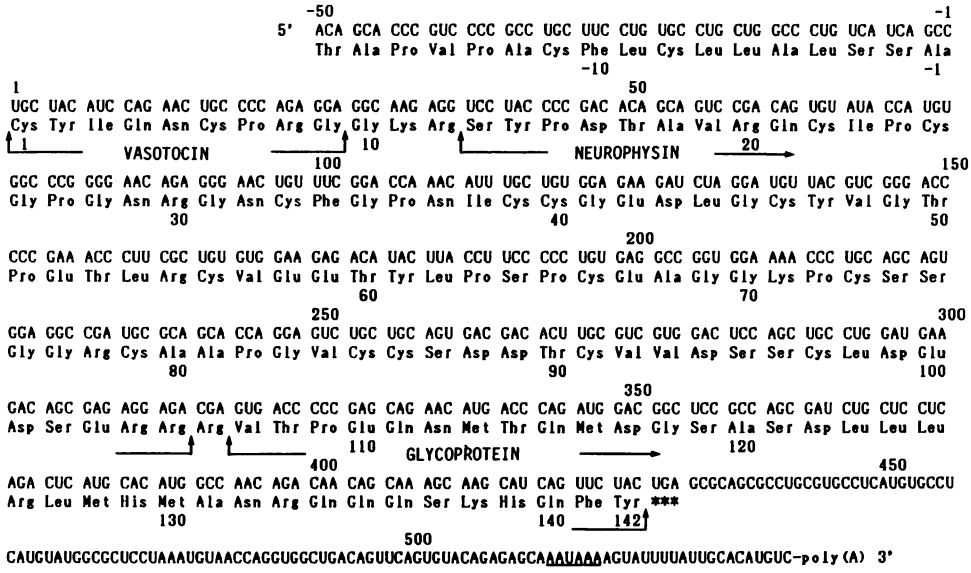

FIG. 4. Primary structure of the mRNA for the VT precursor from the toad brain and the deduced amino acid sequence. Nucleotide and amino acid residues are numbered as shown in Fig. 3.

VT Precursor. The insert of pVT-4 encodes VT at nucleotide positions 1–27. VT is connected to a neurophysin-like peptide by a Gly-Lys-Arg sequence. The carboxyl-terminal region of the VT precursor includes a putative glycoprotein of 36 amino acids (at nucleotide positions 319–426) that is considerably homologous to the [Arg^8]VP glycoprotein and contains the typical *N*-glycosylation site Asn-Xaa-Thr at nucleotides 334–342. This putative glycoprotein is separated from the neurophysin-like peptide by triple arginine residues at nucleotides 310–318. The cloned cDNA for the VT precursor ending at residue −51 does not contain the initiation codon encoding methionine (Fig. 4). Our effort to characterize the missing 5′ nucleotide sequence by using a primer extension technique (19) was unsuccessful, probably because of the occurrence of a CtG-rich region. Thus, the sequence could be deduced for only 17 amino acids of the signal peptide. The amino acids are hydrophobic and homologous to the signal sequence of the MT precursor. The poly(A) tail is preceded by a polyadenylylation signal (AAUAAA) at positions 515–520.

Blot-Hybridization Analysis. The analysis of the toad brain poly(A)$^+$ RNA by RNA transfer blotting with the MT probe revealed only one band of 650 bases (Fig. 5, lane 1), which is consistent with the 600-base mRNA sequence shown in Fig. 3. The RNA blotting analysis of the same preparation with the VT probe showed one band of about 850 bases (Fig. 5, lane 5). This is longer than the poly(A)$^+$ RNA hybridized with the MT cDNA probe. Assuming a length of 100 nucleotides for the poly(A) tail, the VT cDNA clone presumably lacks about 100 nucleotides from the 5′ end of the mRNA. Poly(A)$^+$ RNAs isolated from the testes, ovaries, and livers of toads did not hybridize with either the MT or VT probes (Fig. 5: lanes 2–4, MT; lanes 6–8, VT). This fact suggests that the mRNAs for MT and VT precursors are not transcribed in these tissues, although the possibility that the contents of MT and VT mRNAs are too low to be detected by the present method cannot be excluded.

DISCUSSION

In this report, we have described the cloning and characterization of the cDNAs for the mRNAs encoding the precursors of toad neurohypophysial hormones. The primary structures of MT and VT precursors were deduced from their nucleotide sequences. Our results show that these precursors consist of the signal peptides, hormones followed by Gly-Lys-Arg, and cysteine-rich polypeptides of about 90 amino acid residues that are highly homologous to the OT- and [Arg^8]VP-specific neurophysins (5, 6, 8). The cysteine residues in the molecules of the neurophysins seem to be important for the maintenance of their conformations and functions. The neurophysin-like peptide in the VT precursor is followed by a 36-amino acid peptide including Asn-Xaa-Thr (20), a *N*-glycosylation site with an amino acid sequence analogous to the mammalian [Arg^8]VP-unique glycoprotein (3) (Fig. 6). The toad peptide is somewhat shorter than the mammalian one.

The structural organizations of the MT and VT precursors are thus highly homologous to those of the OT and [Arg^8]VP precursors, respectively. Neurohypophysial nonapeptide hormones are classified into OT-like and [Arg^8]VP-like principles according to their chemical properties and physiological activities (3). Every vertebrate species except cyclostomes has at least one OT-like and one [Arg^8]VP-like principle. It is possible that OT-like and [Arg^8]VP-like families emerged from a common ancestor molecule by gene

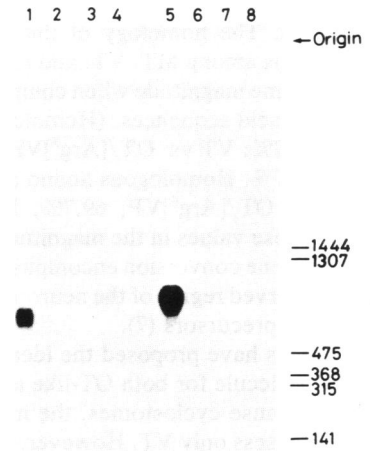

FIG. 5. RNA transfer-blot analysis for MT (lanes 1–4) and VT (lanes 5–8). Two micrograms of the toad brain poly(A)$^+$ RNAs (lanes 1 and 5) and 5 μg of poly(A)$^+$ RNAs from the testes (lanes 2 and 6), ovaries (lanes 3 and 7), and liver (lanes 4 and 8) were electrophoresed and hybridized as described in the text. ^{32}P-labeled pBR322 plasmid digested with *Taq* I was used as a size marker (shown in nucleotides).

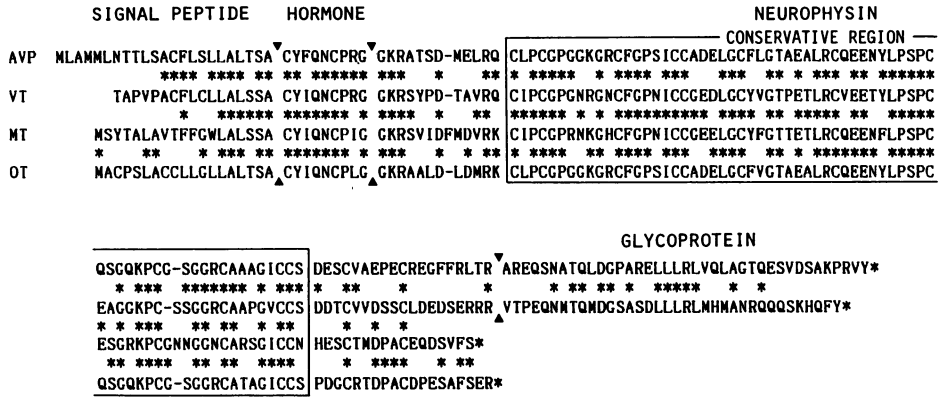

FIG. 6. Comparison of the amino acid sequences among the toad VT and MT precursors and the rat [Arg^8]VP and OT precursors. Gaps indicated by hyphens are introduced to optimize homology. Identical amino acid residues are indicated by asterisks. The conservative regions of neurophysins are indicated by the frame. Arrow heads indicate the boundary between the domains constituting the precursors.

duplication about 450 Myr ago and then evolved independently (21). Amphibian MT has been supposed to be derived from bony fish isotocin and to have been replaced by OT in mammals, while VT was replaced by [Arg^8]VP. The present study strongly supports the hypothesis that VT is the ancestor molecule of [Arg^8]VP, while MT is that of OT. Comparison of the amino acid sequences of their precursors (Fig. 6) supports this hypothesis. The homology between VT and [Arg^8]VP signal peptides is higher than that between VT and MT ones, and MT signal peptide is more homologous to that of OT than VT. This is also true for the carboxyl termini of MT and OT precursors, the homology of which is also higher than that between VT and MT. The amino-terminal peptides in the variable regions of toad VT- and avian VT-specific neurophysins (22) and those of toad MT- and avian MT-specific neurophysins (22), respectively, are considerably homologous. The avian neurophysin amino-terminal peptides, especially of MT, match well with the mammalian counterparts (22). These facts further support the above possibility.

The central regions of OT and [Arg^8]VP neurophysins contain conservative amino acid sequences, regardless of the species from which they were obtained (3). The comparable regions of MT- (at amino acid positions 23–89) and VT-specific neurophysins (at positions 22–87) are also highly homologous (72.7%). The comparable regions of OT- and AVP-specific neurophysins show much higher homology (97%) in rats (Fig. 6). The homology of the conservative regions of neurophysins among MT, VT, and rat OT/[Arg^8]-VP are at nearly the same magnitude when compared by both nucleotide and amino acid sequences. (Homologous nucleotides: VT vs. MT, 67.7%; VT vs. OT/[Arg^8]VP, 67.7%; MT vs. OT/[Arg^8]VP, 70.7%. Homologous amino acids: VT vs. MT, 72.7%; VT vs. OT/[Arg^8]VP, 69.7%; MT vs. OT/[Arg^8]VP, 71.2%.) These values in the magnitudes of homology indicate a recent gene conversion encompassing the exon that encodes the conserved region of the neurophysins within the OT and [Arg^8]VP precursors (7).

Several investigators have proposed the idea that VT is a common ancestral molecule for both OT-like and [Arg^8]VP-like principles (4) because cyclostomes, the most primitive living vertebrates, possess only VT. However, immunocytochemical studies have suggested a presence of OT-like and [Arg^8]VP-like peptides in the nervous system of invertebrates (23). Further studies on the gene structures of these peptides in other species will elucidate the evolutionary pathway of neurohypophysial hormones.

We thank Prof. S. Nakanishi for providing his laboratory manual for the Okayama–Berg method and Drs. A. Kuroiwa and S. Ohsako in the laboratory of T.D. for their helpful discussion. Kent Ross helped us with editing. This study was supported in part by research grants from the Ministry of Education, Science and Culture of Japan.

1. Sawyer, W. H. & Pang, P. K. T. (1979) in *Hormones and Evolution*, ed. Barrington, E. J. W. (Academic, London), pp. 493–523.
2. Jokura, Y. & Urano, A. (1985) *Gen. Comp. Endocrinol.* **60,** 390–397.
3. Acher, R. (1985) in *Neurosecretion and the Biology of Neuropeptides*, eds. Kobayashi, H., Bern, H. A. & Urano, A. (Japan Sci. Soc. Press, Tokyo/Springer, Berlin), pp. 11–25.
4. Gorbman, A., Dickhoff, W. W., Vigna, S. R., Clark, N. B. & Ralph, C. L. (1983) *Comparative Endocrinology* (Wiley, New York), pp. 95–116.
5. Land, H., Schutz, G., Schmale, H. & Richter, D. (1982) *Nature (London)* **295,** 299–303.
6. Land, H., Grez, M., Ruppert, S., Schmale, H., Rehbein, M., Richter, D. & Schutz, G. (1983) *Nature (London)* **302,** 342–344.
7. Ruppert, S., Scherer, G. & Schutz, G. (1984) *Nature (London)* **308,** 554–557.
8. Ivell, R. & Richter, D. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2006–2010.
9. Sausville, E., Carney, D. & Battey, J. (1985) *J. Biol. Chem.* **260,** 10236–10241.
10. Ishida, I., Ichikawa, T. & Deguchi, T. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 308–312.
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
12. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
13. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2,** 161–170.
14. McBride, L. J. & Caruthers, M. H. (1983) *Tetrahedron Lett.* **24,** 245–248.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 202–203.
17. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
18. Brownstein, M. J., Russell, J. T. & Gainer, H. (1980) *Science* **207,** 373–378.
19. Agarwal, K. L., Brunstedt, J. & Noyes, B. E. (1981) *J. Biol. Chem.* **256,** 1023–1028.
20. Pless, D. D. & Lennarz, W. J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 134–138.
21. Acher, R. (1980) *Proc. R. Soc. London Ser. B* **210,** 21–43.
22. Chauvet, M.-T., Lévy, B., Michel, G., Chauvet, J. & Acher, R. (1986) *Biosci. Rep.* **6,** 381–385.
23. Boer, H. H. & Roubos, E. W. (1985) in *Neurosecretion and the Biology of Neuropeptides*, eds. Kobayashi, H., Bern, H. A. & Urano, A. (Japan Sci. Soc. Press, Tokyo/Springer, Berlin), pp. 361–369.