ZOOLOGICAL SCIENCE 9: 157–167 (1992)

Cloning and Sequence Analyses of Vasotocin and Isotocin Precursor cDNAs in the Masu Salmon, *Oncorhynchus masou*: Evolution of Neurohypophysial Hormone Precursors

MASAKAZU SUZUKI, SUSUMU HYODO[1] and AKIHISA URANO

Laboratory of Molecular Biology, Ocean Research Institute
University of Tokyo, Minamidai, Nakano-ku, Tokyo 164, Japan

ABSTRACT—We have cloned and determined the nucleotide sequences of cDNAs encoding precursors of neurohypophysial hormones, vasotocin (VT) and isotocin (IT), from the hypothalamus of masu salmon, *Oncorhynchus masou*. The deduced amino acid sequences of masu salmon VT and IT precursors (proVT-I and proIT-I) are highly homologous to those of chum salmon proVT-I and proIT-I, respectively. The VT and IT precursors are composed of a signal peptide, hormone and neurophysin (NP), the middle portion of which is highly conserved among vertebrates. Both the NPs extend about 30 amino acids at the C-terminal. The extended C-terminals have a leucin-rich segment in the carboxyl-terminal, as copeptin of vasopressin precursor. Southern bot analysis showed the presence of two types of proVT genes (proVT-I and proVT-II) and proIT genes (proIT-I and proIT-II) in an individual masu salmon, as in a chum salmon. Southern blot analysis with proVT probes further suggested that at least two different types of proVT-I genes exist in a single masu salmon. Norhtern blot analysis indicated that proVT-I and proIT-I genes are expressed in the hypothalamus, whereas proVT-II and proIT-II genes are not expressed. Evolutionary distance between proVT-I and proIT-I genes was statistically estimated based on synonymous nucleotide substitution in the coding region of the cDNAs. The magnitude of distance between masu salmon proVT-I and proIT-I genes suggested that the highly conserved central portion of NPs resulted from a gene conversion event. Between masu salmon and chum salmon, evolutionary distance for proVT-I genes is about 6-fold larger than that for proIT-I genes.

INTRODUCTION

Ten distinct neurohypophysial hormones have been characterized in a wide variety of vertebrates, so that many schemes have been proposed for the evolutionary pathway of amino acid substitution based on the primary structures and phyletic distributions of hormones [1, 2]. Meanwhile, a recent molecular biological study clarified that neurohypophysial hormones were synthesized via larger precursors [3]. This fact strongly suggests that it is indispensable to reestimate molecular evolution of neurohypophysial hormones in terms of their precursors.

Accepted August 20, 1991
Received June 19, 1991

[1] Present address: Department of Biology, College of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

The nucleotide sequences of cDNAs encoding the precursors of arginine vasopression (VP) and oxytocin (OT) were first determined in bovine [4, 5]. The deduced amino acid sequence of VP precursor consists of a signal peptide, VP, neurophysin II which is a carrier protein specific to VP, and a glycoprotein named copeptin. The OT precursor is shorter and consists only of a signal peptide, OT and neurophysin I which is OT specific carrier protein. Thereafter, primary structures of neurohypophysial hormone precursors were revealed not only in mammals, e.g. human [6], but also in lower vertebrates such as toad [7], white sucker [8–10] and chum salmon [11, 12]. From these results, Hyodo *et al.* [12] estimated evolutionary distances among mRNAs for precursors on the basis of the nucleotide sequences, and proposed a scheme for molecular evolution of neurohpophysial hormone precursors.

Most salmonids are, like catostomids, considered to be tetraploid, because their cell nuclei contain about 2-fold amount of DNA and chromosomes compared to other species in the same order, Clupeiformes [13]. Furthermore, two different cDNAs could be obtained for each of VT and IT precursors in the white sucker and the chum salmon. Chum salmon precursors were named as provasotocin-I (proVT-I) and II (proVT-II), and proisotocin-I (proIT-I) and II (proIT-II) [12]. Pairs of genes encoding certain hormones were also found in salmonid fish, *e.g.* two genes for melanin-concentrating hormone in the chum salmon [14]. The most important mechanism for generating such pairs of genes is probably genome duplication. Thereafter, the duplicate genes may diverge and individually differentiate. However, there is no detailed study at the nucleotide level about changes of duplicate genes formed by genome duplication. Investigation of the divergence and differentiation of salmonid proVT and proIT genes is thus important to clarify evolutionary changes of neurohypophysial hormone precursor genes in the tetraploid fish. It is also indispensable for general understanding of gene evolution. In the present study, masu salmon was used as an experimental species to clarify the extent of divergence and differentiation within the same genus. We could clone and determine only single proVT-I and proIT-I cDNAs from the hypothalamus of masu salmon (*Oncorhynchus masou*). Northern blot analysis demonstrated that the genes for proVT-II and proIT-II were not expressed, although Southern blot analysis showed the presence of genes for proVT-II and proIT-II.

MATERIALS AND METHODS

Construction of cDNA library, cloning and sequence analysis

The hypothalami were collected from 27 mature female masu salmons. Total RNA was extracted with a RNA extraction kit (Amersham), and poly(A)$^+$RNA was separated through an oligo(dT)-cellulose column using a mRNA purification kit (Pharmacia). Complementary DNA was prepared with a cDNA synthesis system plus

```
a)vasotocin

 amino acids          Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Arg-Gly

                             A
 mRNA            5'-UGC-UAC-AUC-CAA-AAC-UG -3'
                      U   U   U   G   U

                             T
 oligo-VT        3'-ACG-ATG-TAG-GTT-TTG-AC -5'
                      A   A   A   C   A

b)isotocin

 amino acids          Cys-Tyr-Ile-Ser-Asn-Cys-Pro-Ile-Gly

 codons: csIT-I       TGC-TAC-ATC-TCC-AAC-TGC-CCC-ATA-GG

         csIT-II      TGC-TAC-ATC-TCC-AAC-TGT-CCC-ATC-GG

 oligo-IT        5'-TGC-TAC-ATC-TCC-AAC-TGT-CCC-ATA-GG -3'
                                    G           G   C
```

FIG. 1. Synthetic oligonucleotides for the screening of VT and IT coding sequences. a) Oligo-VT is a mixture of 48 oligonucleotides (17-mer) which are complementary to all putative mRNA sequences predicted from the amino acid sequence of VT (1–6). b) Oligo-IT is a mixture of 8 oligonucleotides (26-mer) which were synthesized by considering codon usages in the IT region of chum salmon proIT cDNAs.

(Amersham). A cDNA library was constructed from 50 ng of cDNA and 500 ng of λgt10 arms using a cDNA cloning system λgt10 (Amersham). The cDNA library yielded 1.1×10^5 plaques of 99% recombinants.

A total of 9.8×10^5 transformants was screened by plaque hybridizations. Probes were (1) three cDNAs encoding chum salmon provasotocin-I (cs-proVT-I), provasotocin-II (cs-proVT-II) and proisotocin-II (cs-proIT-II); and (2) two synthetic oligonucleotide mixtures designated as oligo-VT and oligo-IT (Fig. 1). Oligo-VT contained all nucleotide sequences which are complementary to putative mRNAs predicted from the amino acid sequence of VT (1–6), while oligo-IT contained consensus sequences from IT region of chum salmon proIT cDNAs. The cDNA probes were synthesized by a random priming method using a multiprime DNA labeling system and [α-^{32}P]dATP (Amersham). The oligonucleotide probes were 3′-end-labeled with an oligonucleotide 3′-end labeling system (DuPont/NEN) and [α-^{32}P]ddATP (Amersham). Hybridization solution contains 6× SSC (1×SSC contains 150 mM NaCl and 15 mM sodium citrate), 0.1% SDS, 1×Denhardt's reagent, and 100 μg/ml denatured, sonicated calf thymus DNA. Hybridization was performed at 55°C for the cDNA probes, at 35°C for the oligo-VT probe and 50°C for the oligo-IT probe. Post-hybridization washing was carried out twice in 1× SSC containing 0.1% SDS at 50°C for the cDNA probes, while filters were washed twice in 6×SSC containing 0.1% SDS at 35°C for the oligo-VT probe and at 50°C for the oligo-IT probe.

Inserts from positive clones were subcloned into Bluescript plasmid (stratagene), and then nucleotide sequences were determined by the dideoxy chain-termination method [15]. Nucleotide and amino acid sequences were compared by use of a GENETYX genetic information processing software package (Software Development Co., Ltd.).

Southern and Northern blot analyses

Genomic DNA of masu salmon was extracted from a single salmon liver as described by Maniatis *et al.* [16], and was digested separately with each of restriction enzymes, *Eco*R I, *Hin*d III and *Pst* I. The resulting fragments electrophoresed through a 0.6% agarose gel and then transferred to a Hybond-N membrane (Amersham) according to the manufacturer's instructions. Total RNAs (20 μg) extracted from the brain, liver, and kidney of masu salmon were electrophoresed through a 1% agarose/formaldehyde gel, and also transferred to a Hybond-N membrane. The method for labeling cDNA probes and the conditions of hybridization were the same as those in screening procedure with the cDNA probes, except that cDNA probes were generated from cDNAs encoding masu salmon proVT-I (ms-proVT-I), proIT-I (ms-proIT-I), chum salmon proVT-I (cs-proVT-I) and II (cs-proVT-II), and proIT-I (cs-proIT-I) and II (cs-proIT-II). After hybridization, the filters were washed twice in 1×SSC cotaining 0.1% SDS at 60°C, and were exposed to x-ray film at −80°C. Thereafter, the filters were washed twice in more stringent condition, that is, in 0.1×SSC containing 0.1% SDS at 60°C, and again subjected to autoradiography.

Estimation of evolutionary relationship

Evolutionary relationshps among neurohypophysial hormone precursor genes were statistically estimated by the method of Miyata *et al* [17] using Genetyx, genetic information processing software (Software Development Co. Ltd.). The number of substitutions per nucleotide site was calculated in terms of synonymous (Ks) and non-synonmous substitution (Kn) in the coding region.

We first compared masu salmon proVT-I with proIT-I cDNAs. The presence of three exons in proVP and proOT genes of mammals [3] and in proVT genes of white sucker [10] was considered for the calculation, so that the nucleotide sequences of masu salmon proVT-I and proIT-I cDNAs were divided into three regions: (1) region A encoding a signal peptide, hormone and amino (N)-terminal portion of neurophysin (top segment in Fig. 3); (2) region B encoding the central portion of neurophysin (middle segment); and (3) region C encoding the carboxyl (C)-terminal portion of neurophysin (bottom segment) [12]. Then, we compared proVT-I cDNAs and proIT-I cDNAs among masu salmon (ms), chum salmon (cs) and white sucker (ws). Ks was corrected (^{c}Ks) for the

effect of multiple hits at a single site. Evolutionary distance, represented by ^{c}Ks, is directly proportionate to divergence time between homologous genes and also between speices that are to be compared. Therefore, the rate of synonymous substitution per site per year (v) was estimated by the following formula:

$$v=(^{c}Ks_{vt}+^{c}Ks_{it})/4T=(^{c}Ks_{vt}+^{c}Ks_{it})/(4\times1.0\times10^{8})$$

where $^{c}Ks_{vt}$ is ^{c}Ks between ms- and ws-proVT-I cDNAs, $^{c}Ks_{it}$ is ^{c}Ks between ms- and ws-proIT-I cDNAs, and T is divergence time between masu salmon and white sucker. In this study, we adopted 1.0×10^{8} years as T according to fossil record [18] and an isozyme study [19]. By use of the v, we estimated divergence time for proVT-I genes and for proIT-I genes, between masu salmon and chum salmon.

RESULTS

Cloning of cDNAs encoding VT and IT precursors

We obtained 6 positive clones by screening 5.8×10^{5} transformants with the ^{32}P-labeled cDNAs for cs-proVT-I and cs-proVT-II. The analyses of nucleotide sequences of the positive clones showed that 3 clones contained an identical VT-specific seqeunce, while 1 clone contained an IT-specific sequence. These clones were designated as ms-proVT-I and ms-proIT-I cDNAs, respectively, because the homology of the proVT clones to cs-proVT-I cDNA (80.5% nucleotide sequence identity) is higher than that to cs-proVT-II cDNA (48.9%), and the homology of the proIT-I clone to cs-proIT-I cDNA (97.6%) is higher than that to cs-proIT-II cDNA (62.6%). The nucleotide sequences of ms-proVT-I and ms-proIT-I cDNAs and the deduced amino acid sequences are shown in Fig. 2. Since the presence of pairs of cDNAs for both VT and IT precursors were expected, 4.0×10^{5} transformants were rescreened to obtain ms-proVT-II and ms-proIT-II clones with oligo-VT, oligo-IT, and cDNAs for cs-proVT-II and cs-proIT-II. Nevertheless, the sequence of interest was not found in any clones, although we obtained additional two ms-proVT-I and one ms-proIT-I clones.

Vasotocin precursor

The ms-proVT-I is composed of 155 amino acid residues and contains a signal peptide, VT and a neurophysin (VT-NP) which is connected to the hormone by Gly-Lys-Arg sequence. The signal peptide contains a high proportion of hydrophobic amino acids. The Gly-Lys-Arg that follows VT may serve as a signal for proteolytic processing and C-terminal amidation [20]. The VT-NP is cysteine-rich and contains a highly conserved portion at its center, while the C-terminal of VT-NP includes a leucin-rich core segment and shows remarkable similarity to C-terminal of amphibian VT neurophysin and copeptin of mammalian VP precursors, except for a lack of glycosylation site (Fig. 3). The arginine residue at position 103 may be involved in the processing of the precursor because in most mammals the homologous arginine residue in vasopressin precursor is a processing signal between neurophysin and copeptin, although this is not true of amphibian VT precursors [21, 22].

Isotocin precursor

The ms-proIT-I consists of 159 amino acid residues. We predicted that the initiation site for translation of ms-proIT-I is ATG at positions −66 to −64 rather than at positions −60 to −58, because CATGGCT at position −67 to −61 is the same as the consensus sequence for initiation sites of other eukaryotic genes [3, 23]. The ms-proIT-I contains a signal peptide, IT, and a neurophysin (IT-NP), as ms-proVT-I. Isotocin is also connected to IT-NP by Gly-Lys-Arg sequence. Although the central portion of the neurophysin is

FIG. 2. Nucleotide sequences of cDNAs encoding ms-proVT-I and ms-proIT-I precursors and deduced amino acid sequences of the precursors. Nucleotide sites are numbered in the direction from 5′ to 3′, beginning with the first residue in the coding region for hormones. The amino acid residues are numbered with the first residue (Cys) of hormone as 1. Identical nucleotides and amino acids are indicated by colons and asterisks, respectively. Gaps are indicated by hyphens. The AATAAA sequence in the 3′ untranslated region is underlined.

```
provasotocin-I        GGAGTTCAGTTGTAGCCGTAGCCTACAGTACAAATTGGACGAAGCACTTTAGACGGAACAAG -61
proisotocin-I                                 CAAATTCGCAGATCATCCTCACCAAAGCCTCAACCTCAACAC -67

                                  Signal Peptide
          Met Pro Asp Ser Thr Ile Pro Leu Leu Cys Val Leu Gly Leu Leu Ala Leu Ser Ser  -2
--- --- ATG CCA GAT TCT ACT ATT CCA CTT CTG TGC GTC CTG GGG CTC CTC GCG CTC TCC TCT  -2
ATG GCT ATG TTT GGC ACC TCA GTG TCT GCC CTC TGT CTG CTG TTC CTG CTA TCT GTA TGC ACT  -2
Met Ala  *  Phe Gly Thr Ser Val Ser Ala  *   *  Leu  *  Phe  *   *  Ser Val Cys Thr  -2

    1 ---------- Vasotocin ----------
Ala Cys Tyr Ile Gln Asn Cys Pro Arg Gly Gly Lys Arg Ser Phe Pro Asp Leu Lys  -  Arg  19
GCG TGC TAC ATC CAG AAC TGT CCG CGA GGC GGG AAG CGC TCT TTT CCT GAT CTG AAA --- AGA  57
GCC TGC TAC ATC TCC AAC TGC CCC ATA GGA GGC AAG AGA TCA GCC CTA GCT TTC CCA TCC AGA  60
 *   *   *   * Ser  *   *   *  Ile  *   *   *   *   *  Ala Leu Ala Phe Pro Ser  *   20
    1 ---------- Isotocin ----------
Pro Cys Met Ser Cys Gly Pro Gly Asn Arg Gly Leu Cys Phe Gly Pro Ser Ile Cys Cys Gly  40
CCG TGC ATG TCA TGT GGC CCT GGA AAC CGG GGC CTC TGC TTT GGC CCC AGT ATC TGC TGT GGG 120
AAG TGC ATG TCA TGT GGC CCC GGG GAC AGG GGT CGC TGC TTT GGC CCC AAT ATC TGC TGT GGG 123
Lys  *   *   *   *   *   *   *  Asp  *   *  Arg  *   *   *   *  Asn  *   *   *  Arg  41
Glu Gly Met Gly Cys Tyr Met Gly Ser Pro Glu Ala Ala Ser Cys Val Glu Glu Asn Tyr Leu  61
GAA GGG ATG GGT TGC TAC ATG GGC TCC CCA GAG GCA GCT AGT TGT GTG GAG GAG AAC TAC CTG 183
GAG GGG ATG GGC TGC TAC GTG GGC TCT CCA GAG GCA GCT GGC TGC GTG GAG GAG AAC TAC CTG 186
 *   *   *   *   *   *  Val  *   *   *   *   *   *  Gly  *   *   *   *   *   *   *  62

                                  Neurophysin
Thr Ser Pro Cys Glu Val Gly Gly Arg Val Cys Gly Ser Glu Glu Gly His Cys Ala Ala Pro  82
ACC TCT CCC TGT GAG GTC GGA GGA AGA GTG TGT GGG TCT GAG GAG GGA CAC TGT GCT GCA CCT 246
CCC TCC CCC TGT GAG GTC GGA GGA AGA GTG TGT GGC TCT GAG GAG GGA CGT TGT GCT GCA CCG 249
Pro  *   *   *   *   *   *   *   *   *   *   *   *   *   *   *  Arg  *   *   *   *  83
Gly Val Cys Cys Asp Ala Glu Ser Cys Leu Leu Asp Ser Asp Cys Leu Asp Asp Ser Lys Arg 103
GGG GTG TGC TGT GAT GCT GAG AGT TGT CTA CTG GAC TCA GAC TGC CTA GAC GAC AGT AAA CGT 309
GGG ATC TGC TGT GAC GTG GAG GGT TGT AGT ATT GAC CAA TCC TGT ACT GAG GAG GAT GAA GCT 312
 *  Ile  *   *   *  Val  *  Gly  *  Ser Ile  *  Gln Ser  *  Thr Glu Glu Asp Glu Ala 104
Gln Pro Pro Ser Glu Gln Tyr Ser Ser Leu Met Glu Gly Leu Ala Gly Asp Leu Leu Gln Trp 124
CAG CCA CCC AGT GAG CAG TAC AGT TCC TTG ATG GAA GGT TTG GCA GGA GAT CTG CTG CAG TGG 372
GAA TAC ATA AGC CAA TCA GTG AGC AGT --- --- --- AGC CAT GGC CAT GAT CTG CTG ATG AAG 366
Glu Thr Ile  *  Gln Ser Val  *   *   -   -   -  Ser His Gly His  *   *   *  Met Lys 122
Met Leu His Ala Thr Arg Arg Glu Arg Pro Gln                   end                  135
ATG CTG CAT GCC ACC AGA AGG GAG AGA CCT CAG --- --- --- --- TAA CAAACCACTGGCCAGCCCT 427
CTT CTG AAC ATG ATT AGC CAC ACC CCT CCC CAC AGA GTC CAC AAA TAA AACAGCCTCTAATTCAGGG 433
Leu  *  Asn Met Ile Ser His Thr Pro  *  His Arg Val His Lys end                  137

CACCAAAACACACACCCAGAATAGCACCTAGATCAGTTTCACATGCACTACTACTGCAAAAAACCTACACAGCATACACACAC 510
GAGTTGAAAGACAGTGTAAAAAATCTGTCACATCTCAATGTCAGAGGCATACCTATGTACATACATTTTGTACAGAAGACAGA 516

ATCATACACATACAGCAGGACACAGGAAAGGAAGAGTGGGTTTGCTACATAAGAAAAGCAATCAGCTCTAGTACATTCATATT 593
CATAGATGTAAAGGATTAGTTTTGCTTGTAACTGTTTGTGTAATCGCTTGTGTTTCAAGAGTCAAATAAAGCTTACTGTAC  597

TACTGTGCTTACAGTAGCCTTCAGTAAGTTTTAATTTCTAAGGCTTCAACTCAAATATGCATTGTACAATCCACTAGGGGTGG 676

AAGGAATGTAAATATGTAGCAAATAAATATTTTCTTGCACTATT                                        720
```

```
                                  ┌─ Signal peptide ─┐ ┌Hormone┐      ┌──────────
h-provasopressin (h-proVP)        MPDT-MLPACFPGLLAFSSA CYFQNCPRG GKR AMCDL-ELRQ
t-provasotocin   (t-proVT)           TAPVPACFLCLLALSSA CYIQNCPRG GKR SYPDT-AVRQ
cs-provasotocin-I (cs-proVT-I)    MPYSTFQLLWVLGLLALSSA CYIQNCPEG GKR SFPDL-P-RQ
ms-provasotocin-I (ms-proVT-I)    MPDSTIPLLCVLGLLALSSA CYIQNCPRG GKR SFPDL-K-RP
ms-proisotocin-I  (ms-proIT-I)  MAMFGTSVSALCLLFLLSVCTA CYISNCPIG GKR SALAF-PSRK
cs-proisotocin-I  (cs-proIT-I)  MAMFGTSVSALCLLFLLSVCTA CYISNCPIG GKR SALAF-PSRK
t-promesotocin   (t-proMT)        MSYTAL-AVTFFGWLALSSA CYIQNCPIG GKR SVIDFMDVRK
h-prooxytocin    (h-proOT)        MAGPSL-ACCLLGLLALTSA CYIQNCPLG GKR AAPDL-DVRK
```

```
                                  ──────────── Neurophysin ────────────
                                  ┌──────── Conservative region ────────┐
h-proVP    CLPCGPGGKGRCLGPSICCADELGCFVGTAEALRCQEENYLPSPCQSGQKACGS-GGRCAAFGVCCNDE
t-proVT    CIPCGPGNRGNCFGPNICCGEDLGCYVGTPETLRCVEETYLPSPCEAGGKPCSS-GGRCAAPGVCCSDD
cs-proVT-I CMSCGPGDRGRCFGPNICCGEGMGCYMGSPEAAGCVEENYLPSPCEAGGRVCGS-EGSCAASGVCCDSE
ms-proVT-I CMSCGPGNRGLCFGPSICCGEGMGCYMGSPEAASCVEENYLTSPCEVGGRVCGSEEGHCAAPGVCCDAE
ms-proIT-I CMSCGPGDRGRCFGPNICCGEGMGCYVGSPEAAGCVEENYLPSPCEVGGRVCGSEEGRCAAPGICCDVE
cs-proIT-I CMACGPGDRGRCFGPNICCGEGMGCYVGSPEAAGCVEENYLPSPCEAGGRVCGSEEGRCAAPGICCDVE
t-proMT    CIPCGPRNKGHCFGPNICCGEELGCYFGTTETLRCQEENFLPSPCESGRKPCGNNGGNCARSGICCNHE
h-proOT    CLPCGPGGKGRCFGPNICCAEELGCFVGTAEALRCQEENYLPSPCQSGQKACG-SGGRCA-LGLCCSPD
```

```
                                  ┌──────────┐ ┌──────── Copeptin ────────┐
h-proVP    SCVTEPECREGFHRRA R ASD-RSNATQLDGPAGALLLRLVQLAGAPEPFEPAQPDAY
t-proVT    TCVVDSSCLDEDSERR R VTP-EQNMTQMDGSASDLLLRLMHMANRQQQSKHQFY
cs-proVT-I SCVLDPDCL-EDSK-- R QSPSEQNAALMGGLAGDLL-RILH-ATSRGRPQ
ms-proVT-I SCLLDSDCLD-DSF-- R QPPSEQYSSLMEGLAGDLLQWMLH-ATRRERPQ
ms-proIT-I GCSIDQSCTEEDE--- A EYISQSVSS-SHG--HDLLMKLLNMISHTPPHRVHK
cs-proIT-I GCSIDQSCTEEDE--- A EYISQSVSS-SHG--HDLLMKLLNMISHTPPHRVHK
t-proMT    SCTMDPAC-EQDSVFS
h-proOT    GCHADPAC-DAEATFSQR
```

FIG. 3. Comparison of the amino acid sequences among the precursors of masu salmon VT-I and IT-I, chum salmon VT-I and IT-I, toad VT and MT, human VP and OT. Gaps indicated by hyphens have been introduced to maximize homology. Identical amino acids are indicated by asterisks. The conservative regions of neurophysins are enclosed by a frame. Note that the C-terminals of salmonid VT and IT neurophysins include a leucin-rich core segment.

highly homologous to those of amphibian and mammalian neurophysins, the C-terminal, like those in the chum salmon and the white sucker, extends approximately an additional 30 amino acid residues. As C-terminal of VT-NP, this elongated terminal lacks a glycosylation site but includes a leucin-rich core segment (Fig. 3).

Southern and Norhtern bolt analyses

Genomic DNA obtained from a single masu salmon liver was digested separately with *Eco*R I, *Hind* III and *Pst* I, and then hybridized with ms-proVT-I, cs-proVT-I, cs-proVT-II, ms-proIT-I, cs-proIT-I and cs-proIT-II probes. In the autoradiogram of filters washed in 1×SSC-0.1% SDS, the band length pattern yielded by ms-proVT-I cDNA probe was obviously different from that by cs-proVT-II probe but identical with that obtained by cs-proVT-I probe. However, the intensities of hybridization signals with ms-proVT-I probe were somewhat different from those with cs-proVT-I probe (Fig. 4a). After more stringent washing of the same filters in 0.1×SSC-0.1% SDS, the band pattern by cs-proVT-I probe was different from that by ms-proVT-I probe (Fig. 4b). On the other hand, such difference was not found between the patterns yielded by cs- and ms-proIT-I probes after

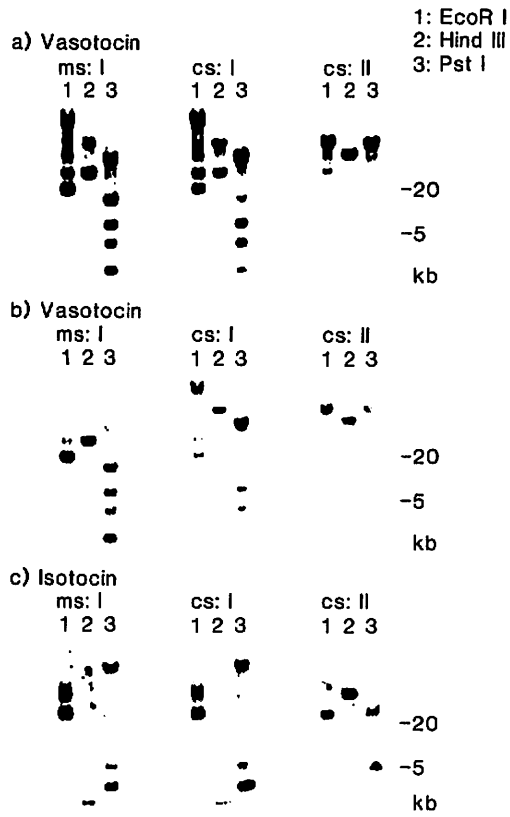

FIG. 4. Southern blot analysis of masu salmon proVT (a, b) and proIT (c) genes in a single masu salmon genome. Five micrograms of DNA from the liver was digested separately with restriction enzymes *Eco*R I (lane 1), *Hind* III (lane 2) and *Pst* I (lane 3) and electrophoresed. Probes used in hybridization are shown above the lane numbers. Filter washing was performed in 1×SSC-0.1% SDS (a) and in 0.1 ×SSC-0.1% SDS (b, c). Positions of size markers, indicated in kb, were obtained by use of *Hind* III and *Eco*R I double digests of lambda DNA.

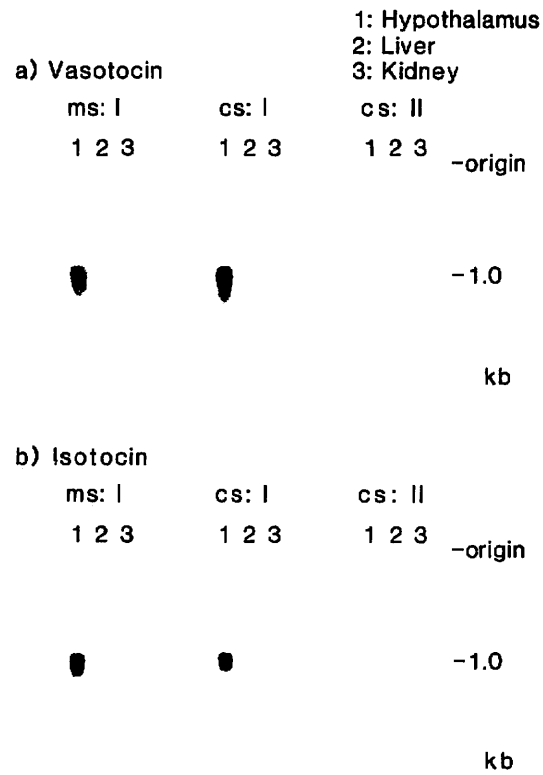

FIG. 5. Northern blot analysis for masu salmon proVT mRNAs (a) and proIT mRNAs (b). Twenty micrograms of total RNAs from the hypothalamus (lane 1), liver (lane 2) and kidney (lane 3) were electrophoresed and hybridized according to the procedure described in the text. Probes used in hybridization are shown above the lane numbers.

filter washing in 0.1×SSC-0.1% SDS, although cs-proIT-II probe detected completely different band pattern (Fig. 4c). These results strongly indicate that, in addition to genes for the analyzed proVT-I and proIT-I mRNAs, an individual masu salmon has proVT-II gene, proIT-II gene and another proVT-I gene which is considerably homologous to cs-proVT-I gene.

Tissue specific expression of ms-proVT and ms-proIT genes was investigated by Northern blot analysis. The ms-proVT-I and cs-proVT-I probes hybridized with a single RNA species of about 900 bases in masu salmon hypothalami, while cs-proVT-II probe did not yield any hybridization signals (Fig. 5a). The proIT probes showed the same trend, i.e. only ms-proIT-I and cs-proIT-I probes detected one band approximately 800 bases in the lane for the hypothalamus (Fig. 5b). RNAs isolated from the kidney and liver of masu salmon did not hybridize with any proVT and proIT probes. These results may explain the fact that we were able to clone ms-proVT-I and ms-proIT-I cDNAs, but unable to obtain ms-proVT-II and ms-proIT-II clones.

Statistical analysis of nucleotide substitution

On the basis of nucleotide substitution, we estimated evolutionary relationships among genes encoding neurohypophysial hormone precursors. Between masu salmon proVT-I and proIT-I cDNAs, the number of substitutions per synonymous site (Ks) and the number of substitutions per nonsynonymous site (Kn) calculated for the central portion of neurophysin (region B) are considerably lower than those for the other regions (Table 1).

Further, we compared proVT-I and proIT-I cDNAs of masu salmon with homologous ones of chum salmon and white sucker, respectively, as is shown in Table 2. Between masu salmon and white sucker, evolutionary distance for proVT-I genes was almost the same as that for proIT-I genes. Therefore, based on these values, we estimated that the mean rate of synonymous substitutions per site per year for proVT and proIT genes was 8.4×10^{-9} in teleosts. Between masu salmon and chum salmon, evolutionary distance for proVT-I genes was about 6-fold larger than that for proIT-I genes. Accordingly, divergence time between ms-proVT-I and cs-proVT-I genes was estimated as 21 myr ago, while that between ms-proIT-I and cs-proIT-I genes as 3.5 myr ago.

TABLE 1. The numbers of synonymous (K_s) and nonsynonymous (K_n) substitutions per nucleotide site for coding region between masu salmon (ms) proVT-I and proIT-I cDNAs. Region A encodes a signal peptide, hormone and N-terminal part of neurophysin; region B, the central portion of neurophysin; and region C, C-terminal part of neurophysin

Sequence compared	K_s	K_n
ms-proVT-I vs. ms-proIT-I		
Region A	0.829	0.343
Region B	0.331	0.060
Region C	0.720	0.528

TABLE 2. Evolutionary distance (^cK_s) and divergence time (T). Interspecies comparison between masu salmon (ms) and chum salmon (cs), and between masu salmon and white sucker (ws)

Sequence compared	^cK_s	T (myr)
ms-proVT-I vs. cs-proVT-I	0.350	21
ms-proIT-I vs. cs-proIT-I	0.058	3.5
ms-proVT-I vs. ws-proVT-I	1.75	100[a]
ms-proIT-I vs. ws-proIT-I	1.61	

[a] This value was estimated according to fossil record [18] and an isozyme study [19].

DISCUSSION

In the present study, the primary structures of masu salmon neurohypophysial hormone precursors, proVT-I and proIT-I, were deduced from the nucleotide sequences of cDNAs. These precursors consist of a signal peptide, hormone and neurophysin (NP). The VT-NP and IT-NP extend about 30 amino acid residues at the C-terminal beyond those described in mammals, so that they are almost of the same length. The central portions of the NPs are particularly similar to each other and the extended C-terminals contain a leucine-rich core segment. Overall high homology between ms-proVT-I and ms-proIT-I supports the

hypothesis that an ancestral molecule diverged into VT and IT, as is the case with neurohypophysial hormone precursors in the chum salmon and the white sucker [9, 12].

We were unable to obtain any clones of masu salmon proVT-II and proIT-II cDNAs. Northern blot analysis did not detect the presence of proVT-II and proIT-II mRNAs. However, Southern blot analysis indicated the presence of proVT-II and proIT-II genes in the masu salmon, as well as in the chum salmon [9, 12] and the white sucker [8, 10].

It is not clear why proVT-II and proIT-II genes were not expressed in the masu salmon, but there are several plausible explanations. Expressions of proVT-II and proIT-II genes may be physiologically repressed in the masu salmon. Alternatively, an occurrence of mutation in 5′-upstream region and/or the coding region may result in the lack of expression. The latter case is supported by the following reports. The gene for vasopressin precursor (proVP) suffered deletion of a single G residue in exon B in the Brattleboro rat [24]. This mutant proVP gene was expressed at a markedly reduced level in the hypothalamus [25]. Furhter, the chum salmon has proVT-II mRNA in which a codon for cystein in proVT-I mRNA is altered to a stop one. The expression level of this mutant proVT-II mRNA in the hypothalamus was considerably low [12].

The comparison of masu salmon proVT-I and proIT-I cDNAs in terms of evolutionary relationship showed that, in region B encoding the highly conserved portion of NP, the number of substitutions per nonsynonymous site (Kn) is considerably smaller than those in the other regions. Furthermore, the number of substitutions per synonymous site (Ks) in region B is much smaller than those in the other regions. These results suggest an occurrence of a gene conversion event encompassing region B of masu salmon proVT-I and proIT-I genes. The possibility of similar gene conversion events was pointed out in the chum salmon and mammals [12, 26, 27]. The exon encoding the central portion of NP may incline to suffer gene conversion irrespective of species.

In this study, we estimated that ms-proVT-I and cs-proVT-I genes diverged about 21 myr ago, although ms-proIT-I and cs-proIT-I genes diverged about 3.5 myr ago (Table 2). An isozyme study suggested that divergence time between masu salmon and chum salmon was about 3.0 myr ago [28]. This divergence time is consistent with that between ms-proIT-I and cs-proIT-I genes, but not with ms-proVT-I and cs-proVT-I genes. It is almost impossible to assume that the rate of molecular evolution of proVT-I genes is very rapid in salmonid, because between masu salmon and white sucker, evolutionary distance for proVT-I genes is comparable to that for proIT-I genes. A possible explanation for the above temporal discrepancy is that proVT-I gene duplicated in a common ancestor of masu salmon and chum salmon. Southern blot analysis showed the presence of two types of proVT-I genes in the masu salmon genome, i.e. the present ms-proVT-I gene and another proVT-I gene which is highly homologous to cs-proVT-I gene. The latter untranscribed proVT-I gene and cs-proVT-I gene may have diverged when the masu salmon and the chum salmon diverged.

In conclusion, the present study suggested that the masu salmon has at least five genes encoding neurohypophysial hormone precursors, among which some regulatory differentiation occurred between proVT-I and II genes, and between proIT-I and II genes. These regulatory differentiation resulted in the lack of expression of proVT-II and proIT-II genes.

ACKNOWLEDGMENTS

We are indebted to Kumagaya Branch of Saitama Prefectural Fisheries Experimental Station for supplying sufficient fresh samples and to Dr. M. Sato, Kyowa Hakko Kogyo, for providing oligo-vasotocin DNA. We express our gratitude to Professor T. Hirano, Ocean Research Institutue, University of Tokyo, for his encouragement and to Mr. S. Sakata, Nippon Oil Co., Ltd., for his advice on various aspects of this study. This study was supported in part by grant-in-aids from the Fisheries Agency, and the Ministry of Education, Culture and Science, Japan.

REFERENCES

1 Gorbman, A., Dickoff, W. W., Vigna, S. R., Clark, N. B. and Ralph, C. L. (1982) Comparative Endo-

crinology. Wiley, New York, pp. 95–116.

2 Acher, R. (1985) Biosynthesis, processing, and evolution of neurohypophysial hormone precursors. In "Neurosecretion and the biology of neuropeptides". Ed. by H. Kobayashi, H. A. Bern, and A. Urano, Japan Sci. Sco. Press, Toky/Springer, Berlin, pp. 11–25.

3 Richter, D. (1987) Oxytocin and vasopressin genes: expression and structure. In "Molecular cloning of hormone genes". Ed. by J. F. Harbener, Humana Press, New Jersey, pp. 173–206.

4 Land, H., Schutz, G., Schmale, H. and Richter, D. (1982) Nucleotide sequence of cloned cDNA encoding bovine arginine vasopressin-neurophysin II precursor. Nature (London), **295**: 299–303.

5 Land, H., Grez, M., Ruppert, S. Schmale, H., Rehbein, M., Richter, D. and Schutz, G. (1983) Deduced amino acid sequence from the bovine oxytocin-neurophysin I precursor cDNA. Nature (London), **302**: 342–344.

6 Sausville, E., Carney, D. and Battey, J. (1985) The human vasopressin gene is linked to the oxytocin gene and is selectively expressed in a cultured lung cancer cell line. J. Biol. Chem., **260**: 10236–10241.

7 Nojiri, H., Ishida, I., Miyashita, E., Sato, M., Urano, A. and Deguchi, T. (1987) Cloning and sequence analysis of cDNAs for neurohypophysial hormones vasotocin and mesotocin for the hypothalamus of toad, *Bufo japonicus*. Proc. Natl. Acad. Sci. USA, **84**: 3043–3046.

8 Figueroa, J., Morley, S. D., Heierhorst, J., Krentler, C., Lederis, K. and Richter, D. (1989) Two isotocin gene are present in the white sucker *Catostomus commersoni* both lacking introns in their protein coding regions. EMBO. J., **8**: 2873–2877.

9 Heierhorst, J., Morley, S. D., Figueroa, J., Krentler, C., Lederis, K. and Richter, D. (1989) Vasotocin and isotocin precursors from the white sucker, *Catostomus commersoni*: Cloning and sequence analysis of the cDNAs. Proc. Natl. Acad. Sci. USA, **86**: 5242–5246.

10 Morley, S. D., Schonrock, C., Heierhorst, J., Figueroa, J., Lederis, K. and Richter, D. (1990) Vasotocin genes of the teleost fish *Catostomus commersoni*: gene structure, exon-intron boundary, and hormone precursor organization. Biochemistry, **29**: 2506–2511.

11 Heierhorst, J., Mahlmann, S., Morley, S. D., Coe, I. R., Sherwood, N. M. and Richter, D. (1990) Molecular cloning of two distinct vasotocin precursor cDNAs from chum salmon (*Oncorhynchu keta*) suggests an ancient gene duplication. FEBS. Lett., **260**: 301–304.

12 Hyodo, S., Kato, Y., Ono, M. and Urano, A. (1991) Cloning and sequence analyses of cDNAs encoding vasotocin and isotocin precursors of chum salmon, *Oncorhynchus keta*: evolutionary relationships of neurohypophysial hormone precursors. J. Comp. Physiol. B., **160**: 601–608.

13 Ohno, S., Wolf, U. and Atkin, N. B. (1968) Evolution from fish to mammals by gene duplication. Hereditas, **59**: 169–187.

14 Takayama, Y., Wada, C., Kawauchi, H. and Ono, M. (1989) Structures of two genes coding for melanin-concentrating hormone of chum salmon. Gene, **80**: 65–73.

15 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, **74**: 5463–5467.

16 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: a laboratory manual. Cold Spring Harbor laboratory, Cold Spring Harbor, New York.

17 Miyata, T., Hayashida, H., Kikuno, R. and Yasunaga, T. (1985) Computer analysis of homology between genes. In "Methods for gene research". Ed. by M. Takanami, S. Nishimura, and M. Matsumura, Tokyo Kagaku Dojin, Tokyo, pp. 381–425.

18 Harland, W. B., Holland, C. H., House, M. R., Hughes, N. F., Reynolds, A. B., Rudwick, M. J. S., Satterthwaite, G. E., Tarlo, L. B. H. and Willey, E. C. (1967) The Fossil Record: a symposium with documentation. Geological Society of London, London, pp. 655–661.

19 Lim, S. T., Kay, R. M. and Bailey, G. S. (1975) Lactate dehyrogenase isozymes of salmonid fish. J. Bio. Chem, **250**: 1790–1800.

20 Brownstein, M. J., Russell, J. T. and Gainer, H. (1980) Synthesis, Transport, and Release of Posterior Pituitary Hormones. Science, **207**: 373–378.

21 Michel, G., Chauvet, J., Chauvet, M. T. and Acher, R. (1987) One-step processing of the amphibian vasotocin precursor: structure of a frog (*Rana esculenta*) "big" neurophysin. Biochem. Biophys. Res. Commun., **149**: 538–544.

22 Chauvet, J., Michel, G., Chauvet, M. T. and Acher, R. (1988) An amphibian two-domain 'big' neurophysin: conformational homology with the mammalian MSEL-neurophysin/copeptin intermediate precursor shown by trypsin-Sepharose proteolysis. FEBS. Lett., **230**: 77–80.

23 Kozak, M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res., **9**: 5233–5252.

24 Schmale, H. and Richter, D. (1984) Single base deletion in the vasopressin gene is the cause of diabetes insipidus in Brattleboro rats. Nature (London), **308**: 705–709.

25 Majzoub, J. A., Pappey, A., Burg, R. and Habener, J. F. (1984) Vasopression gene is expressed at low levels in the hypothalamus of the Brattleboro rat.

Proc. Natl. Acad. Sci. USA, **81**: 5296–5299.

26 Ivell, R. and Richter, D. (1984) Structure and comparison of the oxytocin and vasopressin genes from rat. Proc. Natl. Acad. Sci. USA, **81**: 2006–2010.

27 Ruppert, S., Scherer, G. and Schutz, G. (1984) Recent gene conversion involving bovine vasopressin and oxytocin precursor genes suggested by nucleotide sequence. Nature (London), **308**: 554–557.

28 Numachi, K. (1984) A study on the divergence and phylogeny of salmonids by isozymes. The Heredity (Japan), **38**: 4–11.