Volume 260, number 2, 301-304 FEB 08050 January 1990

Molecular cloning of two distinct vasotocin precursor cDNAs from chum salmon (*Oncorhynchus keta*) suggests an ancient gene duplication

Jörg Heierhorst, Stefan Mahlmann, Steven D. Morley, Imogen R. Coe*, Nancy M. Sherwood* and Dietmar Richter

Institut für Zellbiochemie und Klinische Neurobiologie, UKE, Universität Hamburg, Martinistr. 52, 2000 Hamburg 20, FRG and **Department of Biology, University of Victoria, Victoria, BC, Canada v8w 2Y2*

Received 22 November 1989

The structures of two different vasotocin precursors from chum salmon brain have been elucidated through the molecular cloning of their corresponding cDNAs. Although the predicted precursors, consisting respectively of 153 and 158 amino acids, have the same structural organisation, they show 35% amino acid sequence divergence, of which only approximately half are isofunctional substitutions. Remarkably, while the C terminal segments of both precursors resemble the glycopeptide moiety of the related mammalian vasopressin precursor, both salmon precursors lack consensus sequences for N-glycosylation.

Vasotocin; Neurophysin; Neuropeptide; Gene duplication; Teleost fish

1. INTRODUCTION

The structures of two distinct VT precursor classes from a teleost fish, the white sucker *Catostomus commersoni* have recently been elucidated [1,2]. They show a similar structural organisation to their mammalian (vasopressin) [3] and amphibian [4] counterparts, except that they lack basic amino acid processing signals in positions corresponding to that which separates the neurophysin moiety from the glycopeptide present at the C-terminus of the mammalian vasopressin precursor. They also lack potential consensus sequences for N-glycosylation in their C-terminal regions.

The existence of two distinct VT percursors in the sucker, each encoded by a separate gene [2], is not surprising, since the catostomid fish karyotype probably evolved by tetraploidisation of a cyprinid-like ancestor [5] perhaps 50–100 million years ago. However, the extensive divergence of the sucker VT percursors, amounting to 45% at the amino acid level [2], is not consistent with such an evolutionary time scale, based upon an estimated 1% amino acid divergence rate in 10 million years for related gene pairs [6]. Taken together, these data raise the possibility that sucker VT precursor genes may have been duplicated prior to and independently of tetraploidy.

The salmoniformes separated from the evolutionary line leading to the suckers approximately 70 million years ago [7] and prior to either of these groups attaining a tetraploid karyotype [5]. In order to establish whether the presence of two highly divergent VT-precursors is a general feature of teleost fish, we used a similar approach to that described previously [1] to clone the VT-percursor cDNAs from chum salmon.

Correspondence address: D. Richter, Institut für Zellbiochemie und Klinische Neurobiologie, UKE, Universität Hamburg, Martinistraße 52, 2000 Hamburg 20, FRG

Abbreviations: VT, vasotocin; NP, neurophysin; ar, amino acid residue(s); nr, nucleotide residue(s); SSPE, 0.15 M sodium chloride, 10 mM sodium phosphate (pH 7.4) 1 mM ethylenediaminetetraacetate

These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession numbers X17327 (VT-1) and X17328 (VT-2).

2. MATERIALS AND METHODS

2.1. *cDNA library preparation and screening*

Whole brains were collected from 20 reproductively mature chum salmon that had returned to fresh water to spawn. Fish were obtained from the Qualicum Hatchery on Vancouver Island, BC. cDNA was prepared [8] from 3 μg salmon brain poly(A) + RNA, ligated to *Eco*RI adaptors and inserted into the *Eco*RI site of lambda ZAP-II insertion vector (Stratagene, La Jolla, CA, USA) yielding 1.6×10^6 independent recombinants. After amplification, the library was screened using a fully degenerate 20-mer oligonucleotide pool corresponding to the first 7 amino acids of the mature VT [1]. The oligonucleotide pool, 5'-end-labelled with T4 polynucleotide kinase to a specific activity of 10^8 cpm/μg, was incubated with the replica filters under non-stringent conditions. Subsequently filters were washed stringently at 58°C in the presence of tetramethylammonion ions [9], conditions under which only exact matches to the target sequence should be retained. Positively hybridising lambda ZAP II clones were purified and subjected to helper phage rescue, whereupon the cDNA insert therein was recovered as a subclone in the Bluescript SK(−) phagemid vector [10].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

2.2. *Analysis of cDNA clones*

cDNA inserts or restriction fragments thereof, were sequenced by the dideoxy chain termination method [11], either directly or after subcloning into M13 mp18.

2.3. *Nothern blot analysis*

Poly(A) + RNA (4 μg) was glyoxylated [12], separated in a 1.2% (w/v) agarose gel and transferred to Hybond-N membranes (Amersham) according to the supplier's instructions. Blots were hybridised with ^{32}P-labelled [13] cDNA inserts and finally washed at 65°C with 0.5 × SSPE.

3. RESULTS AND DISCUSSION

Screening of 1.8×10^5 recombinants yielded three positively hybridising clones. Sequence analysis separated these into two classes, designated VT-1 and VT-2 respectively, each of which predicts a distinct chum salmon vasotocin precursor (fig.1). The longest cDNA encoding the VT-1 precursor is 821 nr long, excluding the poly A tail and has in-frame ATG codons at nr 10–12 and 67–69, the latter of which is more likely to be used for translational initiation, since otherwise an atypically long signal peptide would arise. Neither of these start codons is preceded by the consensus sequence for the initiation of translation found in most vertebrate mRNAs [14]. The first stop codon is encountered at nr 526–528, thus giving rise to a predicted VT-1 precursor constituted of 153 amino acid residues and a M_r of 15,997. A classical polyadenylation signal is found at nr 799–804.

In the case of the salmon VT-2 precursor encoded by a cDNA clone of 1199 nr, not including the poly A tail, the first ATG codon is found at nr 45. This is succeeded by an open reading frame of 474 nr predicting a precursor protein of 158 amino acid residues with a molecular mass of 16672 Da. Although a classical polyadenylation signal (AATAAA) is present at nr 1005–1010, the polyadenylation of the VT-2 mRNA seems to be determined, in this case, by the modified sequence ATTAAA present at nr 1174–1179 (fig.1).

Northern blot analysis of the chum salmon poly(A)+RNA used to construct the library, revealed RNA bands of 1100 bases for VT-1 mRNA and 1450 bases for VT-2 mRNA (fig.2). The difference between these values and the lengths of the cDNA clones determined by sequencing is probably accounted for in most part by extensive polyadenylation of the mRNA in vivo. Remarkably, the VT-1 and VT-2 cDNA fragments failed to cross-react under the hybridization conditions used (not shown), emphasising the distinct nature of their sequences.

Both salmon VT-precursors show a similar structural organisation to their mammalian counterpart, the vasopressin precursor [3]. Thus they consist of a putative signal peptide of either 19 (VT-2) or 20 (VT-1)

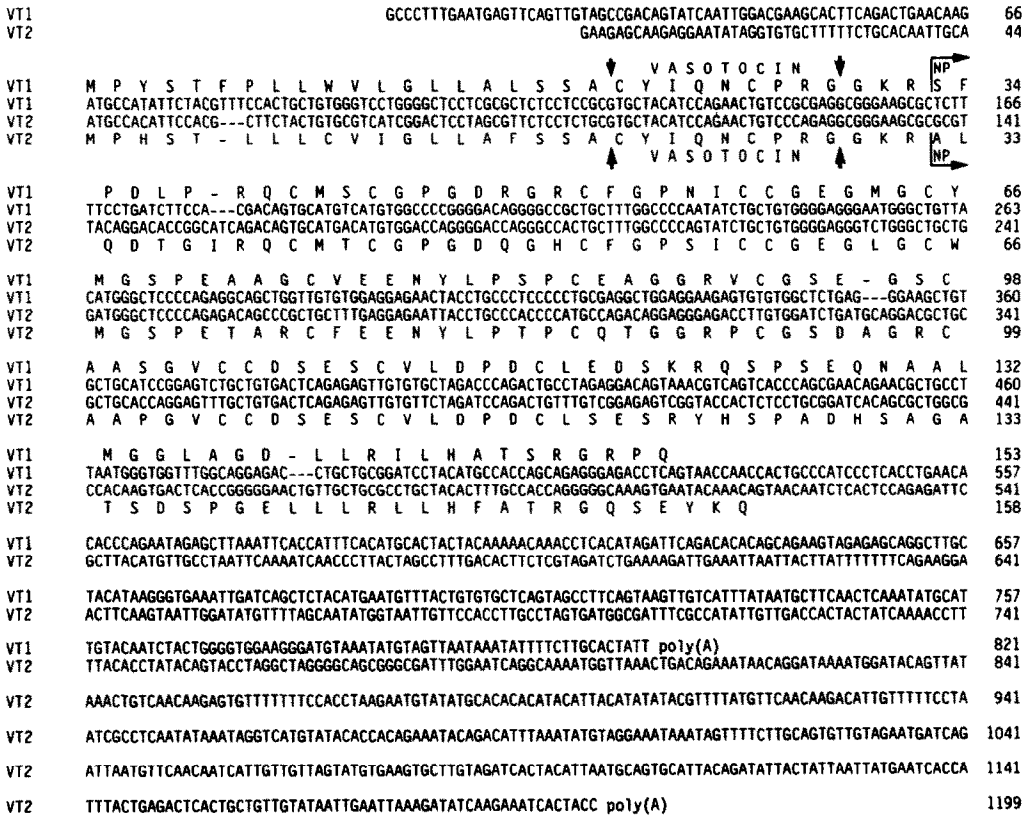

Fig.1. Comparison of the chum salmon VT-1 and VT-2 cDNA sequences and their predicted precursor structures. Hormone sequences (arrowed: ar 21–29; VT-1 or ar 20–28; VT-2) are succeeded by a Gly-Lys-Arg modification and processing sequence and then a neurophysin (NP) moiety. Numbers on the right indicate nucleotide or amino acid positions, respectively.

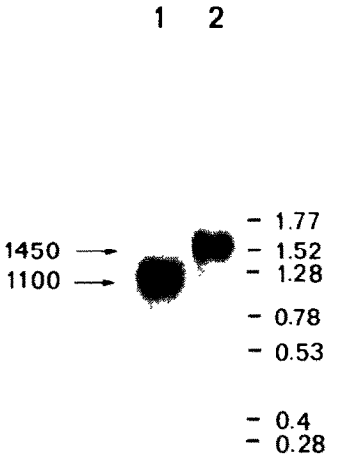

Fig.2. Northern blots of chum salmon poly(A) + RNA. Blots were hybridised with either (lane 1) VT-1 or (lane 2) VT-2 complete cDNA inserts using conditions under which the two cDNAs failed to cross-react. The low molecular weight RNA ladder (BRL) provided molecular size markers indicated on the right in kilobases.

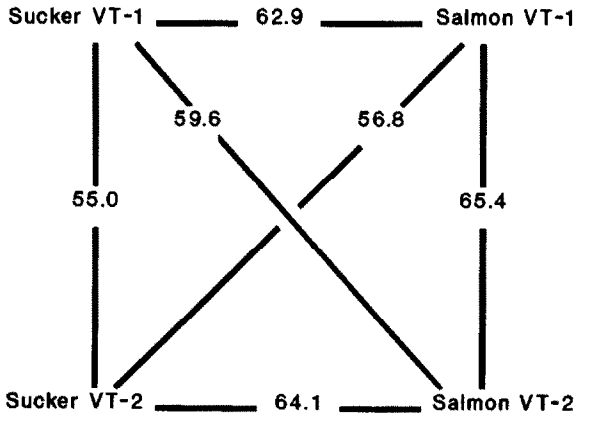

Fig.4. Comparison of the percentage amino acid identity between the salmon and sucker VT-1 and VT-2 percursors. Relationship were assessed using the AALIGN program of the DNASTAR (Madison, WI) sequence analysis software package, and take account of amino acid identities only, after optimal alignment of the two protein sequences.

amino acid residues directly preceding the nonapeptide hormone moiety, which is separated from a cysteine-rich protein resembling the mammalian neurophysin by a modification and processing signal. In each precursor, the neurophysin moiety is succeeded by a tract of approximately 30 amino acids which shows similarities to the mammalian vasopressin-associated glycopeptide, or copeptin [3]. However, both salmon VT precursors lack consensus sequences for N-linked-glycosylation in positions corresponding to that in the copeptin moiety of the vasopressin precursor. A similar phenomenon has recently been demonstrated for the sucker vasotocin- and isotocin-associated neurophysins [1,2] but is in contrast to the situation in all mammalian and amphibian species examined to date [1].

Interestingly, unlike their sucker counterparts, both salmon VT precursors possess putative processing signals (VT-1, ar 121–122; VT-2, ar 122) which could result in the post-translational cleavage of a copeptin-like moiety from the C-terminal end of the salmon neurophysins. However, it has been shown that conformational constraints preclude this second cleavage in an amphibian vasotocin precursor [15]. This has led to the suggestion that the use of such putative cleavage sites occurs only after the transition from vasotocin in non-mammalian vertebrates to vasopressin in mammals [15].

Comparison of the salmon VT precursors with each other and with the sucker VT precursors (fig.3) reveals several important features. Firstly, the salmon VT-1 precursor differs from its relatives by the deletion of a Leu residue from the otherwise extensively conserved leucine-rich core sequence [1]. Secondly, the salmon neurophysins show strict conservation of the 14 cysteine residues found in every neurophysin analysed to date and which form a series of disulphide bridges between different parts of the polypeptide backbone [16]. Nevertheless, the two salmon VT precursors show an overall 35% amino acid divergence of which only half can be accounted for by isofunctional substitutions. This is similar to the situation for the two sucker VT percursors which, however, show an overall 45% amino acid divergence [2].

Thirdly, the salmon VT precursors show considerable amino acid homology to their sucker counterparts (fig.3). This is further emphasised by the fact that, in both cases, the VT-2 percursors include additional

```
      Signal Peptide      Vasotocin  Neurophysin
ccVT1 MSDS-FLPTCILCLLALSSACYIQNCPRGGKRSQPDTS-RECVSCGPGNAGRCYGPSICCGAALGCLVGSPETMSCMEENH
csVT1 MPYSTFPLLWVLGLLALSSACYIQNCPRGGKRSFPDLP-RQCMSCGPGDRGRCFGPNICCGEGMGCYMGSPEAAGCVEENY
csVT2 MPHST-LLLCVIGLLAFSSACYIQNCPRGGKRALQDTGIRQCMTCGPGDQGHCFGPSICCGEGLGCWMGSPETARCFEENY
ccVT2 MSVCAVLLLCVAGLLCLSSACYIQNCPRGGKRALLEPVSRQCLACGPGDKGRCLGPSICCGEEIGCLVGSPWMARCQEEEY

ccVT1 LPSPCETGGRPCGDE-GRCAAPGVCCDSVSCVMDSECLEDVR--SDQSEDPSRLKTVSGEILLRLLNLASRGRRDF
csVT1 LPSPCEAGGRVCGSE-GSCAASGVCCDSESCVLDPDCLEDSKRQSPSEQNAALMGGLAGD-LLRILHATSRGRPQ
csVT2 LPTPCQTGGRPCGSDAGRCAAPGVCCDSESCVLDPDCLSESRYHSPADHSAGATSDSPGELLLRLLHFATRGQSEYKQ
ccVT2 LPSPCQTAGKLCGSDAGPCAAPGVCCGTEGCKLDPNCSEDSESEEPADQNT--LGASPGELLLRLLHPNNRKHNQY
```

Fig.3. Comparison of the predicted VT precursor amino acid sequences of the chum salmon (csVT-1, csVT-2) and the sucker (ccVT-1, ccVT-2) [1,2].

amino acids in identical positions within the first variable (ar 38) and the middle conserved region (ar 96) of their neurophysins, as compared to the VT-1 precursors. The insertion of an extra amino acid between the ninth and tenth cysteine residues of the salmon and sucker VT-2-associated neurophysins is unique among members of the vasopressin precursor family analysed to date. This feature, which is shared only by the toad mesotocin [4] precursor, a member of the oxytocin precursor family, disrupts the otherwise conserved spacing of the cysteine residues, and may perhaps interfere with the folding of the second neurophysin domain [16].

Interspecies similarity exists between the two highly diverged VT precursor types of salmon and sucker and each of the sucker VT precursors shows a greater similarity to its salmon counterpart than to each other (fig.4). As a group the vasotocin precursors are more similar to each other than to any other member of the vasopressin precursor family presently known. The similarities within the group may be a general feature of teleost fish.

Even though the amino acid differences between the chum salmon VT-1 and VT-2 precursors are somewhat lower than that observed for their sucker counterparts, they both suggest the occurrence of an ancestral VT gene duplication substantially predating the separation of the lines leading to the cyprinids and the salmonids, based upon the unit evolutionary period of approximately 10 million years required for the fixation of a 1% amino acid divergence [6]. Since subsequent salmonid evolution after gene duplication resulted in the attainment of a tetraploid karyotype, it remains to be seen if, as proposed for sucker [2], salmon possess four distinct VT percursor genes.

Acknowledgements: We thank Werner Rust and Marion Däumigen-Kullmann (Hamburg) for technical assistance; Dr. Wilhelm Kullmann (Hamburg) for synthesising the oligonucleotides; Carol Warby, David Parker and the staff at Big Qualicum Hatchery for help in collecting the salmon brains. This work was supported by the Deutsche Forschungsgemeinschaft (D.R.) and the Medical Research Council of Canada (N.S.).

REFERENCES

[1] Heierhorst, J., Morley, S.D., Figueroa, J., Krentler, C., Lederis, K. and Richter, D. (1989), Proc. Natl. Acad. Sci. USA 86, 5242–5246.
[2] Morley, S.D., Schönrock, C., Heierhorst, J., Figueroa, J., Lederis, K. and Richter, D. (1989) Biochemistry, in press.
[3] Richter, D. (1987) in: The Peptides (Udenfried, S., Meienhofer, J., eds.) vol. 8, pp. 41–75.
[4] Nojiri, H., Ishida, I., Miyashita, E., Sato, M., Urano, A. and Deguchi, T. (1987) Proc. Natl. Acad. Sci. USA 84, 3043–3046.
[5] Uyeno, T. and Smith, G.R. (1972) Science 125, 644–646.
[6] Wilson, A.C., Carlson, S.S. and White, T.J. (1977) Annu. Rev. Biochem. 46, 573–579.
[7] Carroll, R.L. (1988) Vertebrate Paleontology and Evolution, Freeman, New York.
[8] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[9] Wood, W.I., Gitschier, J., Lasky, L.A. and Lawn, R.M. (1985) Proc. Natl. Acad. Sci. USA 82, 1585–1588.
[10] Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) Nucl. Acids Res. 16, 7583–7600.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Thomas, P.S. (1983) Methods Enzymol. 100, 255–266.
[13] Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266–267.
[14] Kozak, M. (1987) Nucl. Acids Res. 15, 8125–8148.
[15] Chauvet, J., Michel, G., Chauvet, M.-T. and Acher, R. (1988) FEBS Lett. 230, 77–80.
[16] Burman, S., Wellner, D., Chait, B., Chaudhary, T. and Breslow, E. (1989) Proc. Natl. Acad. Sci. USA 86, 429–433.