The Human Vasopressin Gene Is Linked to the Oxytocin Gene and Is Selectively Expressed in a Cultured Lung Cancer Cell Line*

(Received for publication, November 6, 1984)

Edward Sausville, Desmond Carney‡, and James Battey§
From the National Cancer Institute, Navy Medical Oncology Branch, Clinical Oncology Program, Division of Cancer Treatment, National Cancer Institute, Naval Hospital Bethesda, Bethesda, Maryland 20814

The human genes for prepro-arginine-vasopressin-neurophysin II (prepro-AVP-NPII) and prepro-oxytocin-neurophysin I (prepro-OT-NPI) were cloned from a human genomic library and the nucleotide sequence of both genes was determined. The two genes are similar in their intron-exon structure, linked together with 12 kilobases intervening, and transcribed from opposite DNA strands. A human small cell lung cancer cell line, H378, produces significant quantities of prepro-AVP-NPII mRNA using a transcription unit predicted from the genomic DNA sequence. Despite the proximity of the actively transcribed prepro-AVP-NPII gene, transcription of prepro-OT-NPI is not detected in this cell line.

The conventionally accepted activities of the posterior pituitary hormones AVP and OT are in the regulation of water excretion, parturition, and lactation. In addition, AVP receptors and AVP-like immunoreactive material exist in central and autonomic neurons (Hanley et al., 1984; Buijs et al., 1983). AVP is derived from a larger precursor, prepro-AVP-NPII, whose primary nucleotide structure was first elucidated in a bovine cDNA clone by Land et al. (1982). The precursor consists of a single polypeptide containing a secretory signal sequence (Blobel and Dobberstein, 1975), the nonapeptide AVP (preceded by a signal peptide), the hormone carrier protein neurophysin II, and from the carboxyl terminus another peptide of unknown function but resembling a previously detected pituitary glycoprotein. Bovine prepro-OT-NPI has an analogous structure (Land et al., 1983; Ruppert et al., 1984) except that OT is present in place of AVP and neurophysin I is present in place of neurophysin II. There is no peptide in prepro-OT-NPI analogous to the carboxyl-terminal glycoprotein present in prepro-AVP-NPII. Although the primary sequence of prepro-AVP-NPII and prepro-OT-NPI has been characterized in cow and rat, it has not been known whether these two structurally similar genes are linked in the genome.

Human small cell lung cancer (SCLC) is distinguished from other types of lung cancer by selective elaboration of polypeptide hormones, including AVP, pro-opiomelanocortin, calcitonin, and bombesin/gastrin-releasing peptide (Gazdar and Carney, 1984). As a prelude to understanding the basis for selective hormone gene expression by certain SCLC tumors and cell lines we have cloned the human prepro-AVP-NPII gene from a genomic library. The same screening procedure also allowed us to characterize the human prepro-OT-NPI gene. We report here the structure of the human prepro-AVP-NPII and prepro-OT-NPI genes and show that they are physically linked within 12 kb in the human genome with an inverted arrangement of their coding strands. In addition, a SCLC-derived cell line transcribes this prepro-AVP-NPII gene in a way predicted by the gene structure. In contrast, transcription of prepro-OT-NPI in this cell line is not observed.

MATERIALS AND METHODS

RESULTS AND DISCUSSION

Structure of the Human Prepro-AVP-NPII Gene—A probe from the neurophysin portion of bovine prepro-AVP-NPIII was obtained by preparing a 0.3-kb Smal-PouI fragment from a cDNA clone (pVNPII-1) of Land et al. (1982). This probe (designated cc-bov) has homology to both neurophysins I and II, due to their extensive nucleic acid sequence similarity (Land et al., 1983; Schmale et al., 1983). In addition, a 0.2-kb Pstl-Smal fragment from the AVP-specific 5' region of bovine prepro-AVP-NPII, designated 5' bov was also used. Of 10 plates from a human genomic library derived from MboI cut placental DNA cloned into Charon 28A, eight positive clones were obtained and characterized.

The library clones fell into two classes represented by hAVP3 and hAVP4, whose structure is outlined in the top portion of Fig. 1. Homology to both 5' bov and cc-bov exists on the 15-kb insert of hAVP4. Restriction mapping of hAVP4 reveals that homology to cc-bov is found on both 2.5- and 5.0-kb BamHI-BamHI fragments. Homology to the 5' bov probe in hAVP4 exists only on the 2.5-kb fragment. The fact that homology with cc-bov in hAVP3 is linked to phage arms when the phage is digested with BamHI shows that the 2.5-kb BamHI-BamHI fragment in hAVP4 is bounded by an artificial BamHI site on its 5' border that is created from an MboI site.

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

‡ Current address: Mater Misericordiae Hospital, Eccles St., Dublin 7, Ireland.
§ To whom reprint requests should be addressed.
2 The abbreviations used are: AVP, 8-arginine-vasopressin; OT, oxytocin; prepro-AVP-NPII, prepro-AVP-neurophysin II; prepro-OT-NPI, prepro-OT-neurophysin I; kb, kilobase, SCLC, small cell lung cancer; hAVP, human 8-arginine-vasopressin.
The human prepro-AVP-NPII gene has a three-exon genomic structure in common with cow and rat (Land et al., 1982; Ivell and Richter, 1984). The human AVP peptide is derived from a protein precursor comprised of a signal peptide, AVP, neurophysin II, and another peptide in the 3′ untranslated portion of this gene continues through a polyadenylation signal (Proudfoot and Brownlee, 1976) at position 2309, and an approximate 3′ border of the third exon is indicated at about position 2345.

Therefore, the human prepro-AVP-NPII gene is localized to the 2.5-kb BamHI-BamHI fragment of hAVP4, homology to the cc-bov fragment derived from a bovine cDNA clone can also be found on the 5-kb BamHI-BamHI fragment present in hAVP4. The AVP-specific 5′ bov cDNA probe was limited in its homology to the 2.5-kb BamHI-BamHI fragment from hAVP4 (Fig. 1). To reconcile these observations, a 1.4-kb region of hAVP4 containing the additional homology region was sequenced.

The only difference observed between our predicted sequence for mRNA initiation (Breathnach and Chambon, 1981) at position 355, the first ATG thereatere defines an open reading frame containing a 9 amino acid peptide segment expected for human oxytocin (nucleotides 475–501). By analogy to the bovine prepro-OT-NPII gene (Land et al., 1983), the human oxytocin sequence (residues 475–501) is preceded by a largely hydrophobic peptide from residues 418–474, and followed by a tripeptide (Gly-Lys-Arg) before commencing a putative translation product with a sequence expected for human neurophysin I (Chauvet et al., 1983) at position 511. A break in homology with the expected translation product occurs at position 538, with the characteristics of a splice donor site (Breathnach and Chambon, 1981), defining the first exon of human prepro-OT-NPI.

A 299-bp intervening sequence separates the first and second exon, and homology with human neurophysin I resumes (position 838). The second exon is comprised solely of neurophysin 1-related sequences, extending to position 1036. A short intervening sequence (83 bp) separates the second and third exons, with the remaining sequences of neurophysin I encoded in the third exon. As in the cow and rat (Ruppert et al., 1984; Ivell and Richter, 1984), a termination codon follows directly after the neurophysin I coding sequences with no apparent analog of the 3′ putative glycopeptide segment found in prepro-AVP-NPII. A double consensus polyadenylation signal (Proudfoot and Brownlee, 1976) is found at position 1248, indicating the end of the transcription unit.

The only difference observed between our predicted se-
Human Vasopressin and Oxytocin Genes

FIG. 2. Sequence of the human prepro-AVP-NPII gene. The region enclosed by the thin solid black line indicates presumptive exons of the prepro-AVP-NPII gene indicated as ex1, ex2, and ex3, respectively. The heavy black line in exon 1 encloses the 8-arginine-vasopressin nonapeptide. The dashed black line in exons 1, 2, and 3 enclosing stippled pattern indicates sequences found in neurophysin II. The remaining nucleotides in exon 3 comprise the putative human AVP-NPII-associated glycoprotein. The overline indicates a consensus recognition signal for RNA polymerase II and the underline a consensus polyadenylation signal. Below the sequence the strategy for sequencing is indicated, with each arrow head representing a discrete m13 clone, and the extent of nucleotide sequence determined. The 5' and 3' untranslated regions are indicated as open bars in the BstHI-RsaI and PvuII-PstI fragments adjacent to dark bars representing exons 1 and 3, respectively. Restriction endonuclease site abbreviations are as described in the legend to Fig. 1.

Quence for neurophysin I and the amino acid sequence determined by Chauvet et al. (1983) consists of a valine proposed by those workers to exist between what would be the position indicated by nucleotides 1014 and 1015 (Fig. 3). This is not predicted from our nucleotide sequence. This difference may reflect a sequencing error or an allelic difference between the gene producing the neurophysin I sequenced as protein and the gene cloned and sequenced from the library here.

Orientation and Comparison of AVP and OT Genes—The nucleotide sequence of prepro-AVP-NPII and prepro-OT-NPI predicts that the two genes exist in opposite transcriptional orientations. This was confirmed by hybridizing OT-specific and AVP-specific probes to HindIII-SstI digests of hAVP4 blotted to nitrocellulose. An 11.7-kb HindIII-SstI fragment hybridized to exons 2 and 3 of both prepro-AVP-NPII and prepro-OT-NPI, but not to exon 1 of prepro-AVP-NPII (data not shown). Since DNA sequencing establishes the 5' to 3' orientation of prepro-OT-NPI to be from the BamHI site to the SstI site (Fig. 3), the size (11.7 kb) of this HindIII-SstI fragment can only be explained if the two genes are in opposite orientations.

Nucleotide sequences encoding the nonapeptide AVP and OT hormones are highly homologous. In addition, dot matrix analysis (Maizel and Lenk (1981) reveals (data not shown) that the sequence of a portion of the first intron and exon 2 of prepro-AVP-NPII (nucleotides 1512-1920) is highly homologous to part of the first intron and exon 2 of prepro-OT-NPI (nucleotides 620-1036). Analogous features of the bovine genes for prepro-AVP-NPII and prepro-OT-NPI had previously been noted (Ruppert et al., 1984). The results of our
sequence analysis suggest that to explain the evolution of these two structurally similar genes from a common ancestor, both gene duplication and an inversion are required. Apparent duplication and inversion has also been described in the structurally related egg-laying hormone-related merge of *Aplysia* (Scheller et al., 1983), although its significance in this system is unclear. Boorstein et al. (1982) have noted a duplicated and inverted arrangement of genes potentially encoding the B-subunits of human chorionic gonadotropin. Consequences of a inverted arrangement of related genes could include the establishment of differences in transcriptional regulation, chromatin protein binding properties, and decreased recombination frequency.

Extensive interspecies homology is found when the bovine and rat 5' flanking sequences of prepro-AVP-NPII are compared (Ruppert et al., 1984). This is also true for 5' flanking sequences of cow and rat prepro-OT-NPI. The human genes for these two hormones also share this interspecies homology of 5' flanking sequences, although there is negligible homology in 5' flanking sequences when the two human genes are compared to each other (data not shown). Since it has recently been shown that expression of a polypeptide hormone gene can be influenced in a tissue specific fashion by the adjacent 5' flanking sequences (Supowit et al., 1984) it is possible that these conserved sequences are responsible for the differential expression of AVP-NPII and OT-NPI in tissues which specifically produce either hormone. In *vitro* alteration of these flanking sequences followed by DNA transfection and expression studies in appropriate cell hosts may allow sequences important in differential hormone gene expression to be defined.

Acknowledgments—We would like to thank L. Hennighausen for his generous gift of the bovine cDNA clone. A. Wolfson of the UCLA Medical Center determined AVP levels in SCLC lines. In addition, we thank A. Gazdar, J. Minna, M. Kuehl, G. Hollis, T. Bender, and I. Kirsch for helpful discussions, as well as J. Maizel and J. Owen for assistance with the dot matrix analysis of DNA sequences.

REFERENCES


Additional references are found on p. 10241.

Continued on next page.
Supplementary Material to
The Human Vasopressin Gene is Linked to the Oxytocin Gene and is
Selectively Expressed in a Cultured Long-Term Cell Line
by Edward Sausville, Raymond Carney, and James Batten
Materials and Methods
Cloning
A recombinant Charon DNA (Kleber et al., 1977) library of 1 x 10^9 clones was screened as described with a nucleic acid probe derived from phage 7 of Land et al., (1982), which was a gift to J. Batten from L. Henninghausen. Plasmid subclones were in all cases constructed in pBR322, a derivative of pBR327 with the polylinker region of my 11 DNA substituted for the small EcoRI-HindIII fragment. Plasmid growth in general was in E. coli LE392 under Km F+ condition conditions.

DNA Sequencing
M13 mp 10 and mp 11 subclones were derived and sequencing performed essentially as described by the Bethesda Research Laboratory Manual (1985). Where favorable restriction sites did not exist, site-selected exonuclease Bal31 digests of plasmids were cloned into the Smal site of M13 phages. Approximately 90% of the prepro-AVP-NPFI gene and 90% of the prepro-
OFP gene was sequenced from both strands.

Cell Lines
Fresh specimens from patients with SSCC were minced and grown in HITES medium (Bioko et al., 1976) at 37°C, and poly(A)+ RNA was prepared as described by Scriver et al., (1979). A total of 2 x 10^6 cells was used for each preparation. The RNA was separated on agarose gels containing 2.2 M formamide (Ehrlich et al., 1977) and transferred to nitrocellulose. All hybridizations to DNA or RNA immobilized on nitrocellulose, indicated fragments were prepared free of plasmid vector by preparative electrophoresis and then labeled to high specific activity with [32P]dCTP as described by Sambrook et al., (1989). Hybridization to nitrocellulose blots of transferred RNA and DNA in 50% dextran sulfate was conducted as described by Wabue et al., (1982) and used as a probe at 1 x 10^6 cpm/ml. Northern analysis were as described by Latt et al., (1983). Single stranded probes complementary to cDNA sequences were synthesized by primer extension of an m13 subclone in the presence of [32P]dATP.

RESULTS AND DISCUSSION
Organization of AVP and OT Genes
The structure of the prepro-AVP-NPFI gene in SSCC derived from the fetal liver library predicts that 5 exons of DNA the gene should occur on a 4.9 kb BamHI fragment, and both prepro-AVP-NPFI and prepro-DOPFI should be contained on a single Sall fragment. Figure 4 demonstrates that DNA derived from SSCC, a SSCC cell line which produces substantial quantities of immunoreactive AVP and DNA from SSDR, a SSCC cell line not producing AVP, have a similar pattern of hybridization to both AVP and OT specific probes. An AVP exon 1 specific probe hybridized to a 9.6 kb BamHI fragment and a 19 kb Sall fragment. An OT exon 3 specific probe hybridized to the same Sall fragment as well as the expected 4.9 kb BamHI fragment. These results verify the physical linkage of prepro-AVP-NPFI and prepro-DOPFI predicted from the structure of the closed gene. The pattern of hybridization shown in Fig. 4 was identical to that seen in DNA from normal lymphocytes (data not shown), and analogous results with DNA from the other SSCC cell lines suggest no frequent amplification or rearrangement of the prepro-AVP-NPFI gene in SSCC.

Expression of Prepro-AVP-NPFI in SSCC
To validate the predicted transcription unit of prepro-AVP-NPFI in human, the expression of this gene was studied in two SSCC cell lines. SSTV produces substantial quantities of AVP, whereas SSDR does not.

Fig. 5, panel 1 shows that a 1.5 Kb BamHI fragment (probe 1) encoding only the first exon of prepro-AVP-NPFI detects a 0.7 kb species in poly(A) mRNA from SSTV and is not found in RNA from SSDR. In addition, a 1.5 Kb species is also detected by probe 1 (poly(A) mRNA only from SSTV. This material is clearly separated from the observed position of the 1.4 kb ribosomal subunit. Probe 1 also apparently hybridizes cap-specific to the 5' RNA in poly(A) mRNA from normal lymphocytes.

In panel 1, probe 2 shows results with probe 2, deriving from the first interesting sequence downstream of the 5' side of the prepro-AVP-NPFI gene. This probe hybridizes to both the 1.5 kb and 2.5 kb species in mRNA from SSTV but not SSDR. Since this probe contains the neurophysin sequences, it would not be expected to distinguish between AVP and OT transcripts. However, Fig. 5, panel 4 reveals that the OT-specific exon 3 3' end-probe 4 does not specifically hybridize to any poly(A) mRNA from SSTV or SSDR.
Human Vasopressin and Oxytocin Genes

Figure 5. Transcription of Prepro-AVP-NPI in SCLC Cell Lines.

As an implication of the results presented in Fig. 5, it is that there is selective expression of the prepro-AVP-NPI gene but not the prepro-OXT-NPI gene, despite their close proximity.

Figure 6. SI-nucleus Protection assay by RNA from Human SCLC Cell Lines. A single-stranded RNA probe was prepared by primer extension using an oligonucleotide upstream from the prepro-AVP-NPI gene containing the 5'-untranslated portion and the entire first exon. The probe was hybridized to total cellular RNA and poly(A) RNA from the indicated SCLC cell lines. Following hybridization, non-annealed probe was largely digested with SI nuclease, and the products of digestion separated on an 8% gel. Residual unannealed probe (approximately 15% of input probe) is present in all samples at 348 nucleotides.

Reference:
Southern, P.M. (1975) J. Mol. Biol. 98, 503 - 517