The EMBO Journal Vol.2 No.5 pp.763 – 767, 1983

Structural organization of the rat gene for the arginine vasopressin-neurophysin precursor

Hartwig Schmale, Susanne Heinsohn and Dietmar Richter*

Institut für Physiologische Chemie, Abteilung Zellbiochemie, Universität Hamburg, Martinistr. 52, D-2000 Hamburg 20, FRG

Communicated by G. Kreil
Received on 21 February 1983

The rat arginine vasopressin-neurophysin precursor gene has been isolated from a genomic library cloned in λ phage Charon 4A. Restriction mapping and nucleotide sequence analysis demonstrated that the gene is 1.85 kilobase pairs long and contains two intervening sequences located in the protein coding region. Exon A encodes a putative signal peptide, the hormone arginine vasopressin and the variable N terminus of the carrier protein neurophysin, exon B encodes the highly conserved middle part of neurophysin and exon C its variable C terminus together with glycoprotein. Thus, the three functional domains of the precursor – arginine vasopressin, neurophysin, glycoprotein – are encoded on three distinct exons.

Key words: vasopressin gene/polyprotein/gene conversion

Introduction

Arginine vasopressin (AVP) and oxytocin (OT), two nonapeptide hormones, are produced together with their corresponding 'carrier' proteins, the neurophysins, as common precursors by different magnocellular neurones in both the supraoptic and paraventricular nuclei of the hypothalamus (Pickering and Swann, 1983; Brownstein *et al.*, 1980; Breslow, 1979; Acher, 1979; North *et al.*, 1978). The precursors are packaged into neurosecretory granules and transported axonally in the pituitary stalk to the posterior pituitary. En route they are processed into the hormones and the neurophysins (Brownstein *et al.*, 1980). There are two types of neurophysins, associated with either OT or AVP. They consist of 93 – 95 amino acid residues of which a central sequence of 74 amino acid residues is highly conserved (Chauvet *et al.*, 1981).

The primary structures of the bovine AVP and OT preprohormones have recently been elucidated by sequencing the cDNAs encoding the respective hypothalamic mRNAs (Land *et al.*, 1982, 1983). In both precursors, the hormone immediately follows the signal peptide and is adjacent to its respective neurophysin (Np); prepro-AVP-Np additionally includes a glycoprotein at the C terminus. The structure of the rat AVP-Np gene could help in understanding not only the synthesis of AVP but also the molecular mechanism underlying the defect in the expression of AVP in the Brattleboro strain of rats (Valtin *et al.*, 1974). We have now analysed the structure of a genomic DNA fragment encoding the AVP-Np gene from normal rats. Its DNA sequence reflects the organization of the bovine protein precursor (Land *et al.*, 1982) with three functional domains – AVP, Np, glycoprotein – each on a separate exon indicating an organization different from that found in the pro-opiocortin or pro-enkephalin genes (Herbert, 1981; Kakidani *et al.*, 1982; Numa and Nakanishi, 1981; Noda *et al.*, 1982; Gubler *et al.*, 1982; Legon *et al.*, 1982).

*To whom reprint requests should be sent.

Results

A rat genomic DNA library in the lambda vector Charon 4A was screened for bacteriophages carrying AVP-Np gene sequences by hybridization with the ^{32}P-labeled bovine pVNpII-1 cDNA (Land *et al.*, 1982). The AVP-Np gene was identified in the clone λ Charon 4A-VN3 by hybridization to three *AvaI-PstI* restriction fragments of the cDNA encoding the N-terminal, central and C-terminal part of the bovine AVP-Np precursor (Land *et al.*, 1982). Subsequent restriction mapping and blot hybridization (Southern, 1975) of DNA from this clone delimited the AVP-Np gene to a single 3.8-kb *HindIII* restriction fragment. This was subcloned in the plasmid pUC 8 (clone pVλ *HindIII*) and used for more detailed restriction mapping and sequence determination (Maxam and Gilbert, 1980), following the strategy outlined in Figure 1.

Comparison of the rat genomic DNA sequence with the bovine cDNA (Land *et al.*, 1982) encoding the AVP-NpII precursor has enabled us to locate three exons and two intervening sequences (introns) in the rat gene (Figure 2). Exon A encodes a putative signal peptide, the hormone and the variable N terminus of Np, exon B the highly conserved middle part of Np including nearly all cysteine residues essential to secondary structure (Cohen *et al.*, 1979; Breslow, 1979) and exon C the variable C terminus together with the glycoprotein. Thus, the rat DNA sequence predicts a protein precursor similar in structure to the bovine one (Land *et al.*, 1982). The two introns are inserted in the protein coding sequence specifying the rat neurophysin (Chauvet *et al.*, 1981). Intron I (~1 kb) exactly separates the variable N terminus and the constant part of the neurophysin at position 21 of the precursor. The second intron of 227 bp occurs between the first and second nucleotides of the GAG which encodes Glu

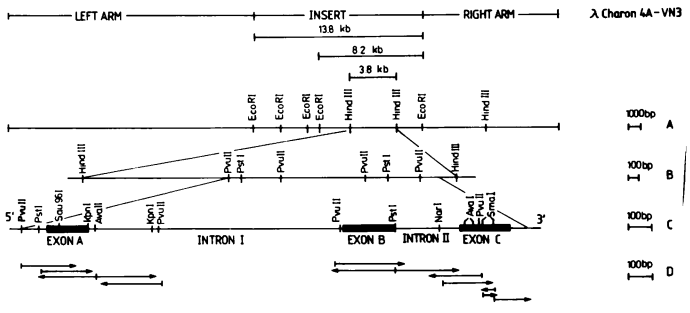

Fig. 1. Restriction mapping and sequencing strategy of cloned genomic DNA fragments containing the AVP-Np gene. **(A)** The clone λ Charon 4A-VN 3 displaying all existing sites for the restriction enzymes indicated. **(B)** Partial restriction map of subclone pVλ*HindIII*. Only the relevant restriction sites are indicated. **(C)** The exon-intron structure of the rate AVP-Np precursor gene. The relative positions of the exons are shown by thick bars. Only relevant sites for restriction enzymes are displayed. **(D)** Sequencing strategy; the horizontal arrows indicate the direction and extent of sequence determination.

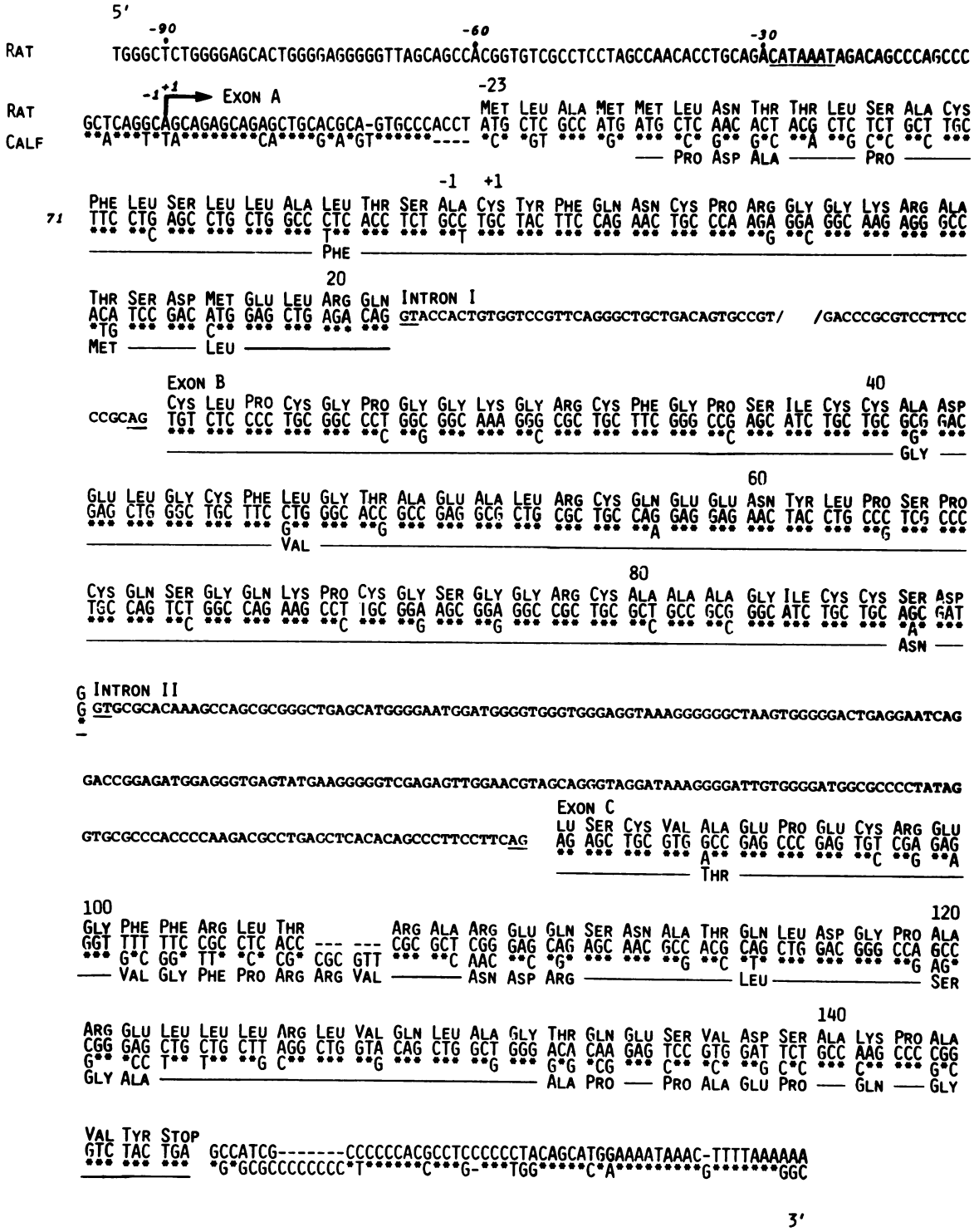

Fig. 2. Sequence of the rat prepro-AVP-neurophysin gene in comparison with that of bovine cDNA. The nucleotide sequence of the rat anti-sense strand is shown together with the deduced amino acid sequence. Nucleotide (asterisks) and amino acid (solid lines) sequence homologies found in the bovine cDNA are indicated beneath the rat sequences. Absence of nucleotides is indicated by dashed lines, absence of amino acids by a gap. No symbols beneath the rat sequences indicate that the respective bovine nucleotides are not yet known. The amino acid residues are numbered, the negative numbers indicate the putative signal sequence of the rat precursor. Italic numbers indicate the nucleotide sequence upstream (negative numbers) and downstream (positive numbers) of the transcription start site. Introns are shown in small letters; the sequence of intron I is only partially shown. The assignment of the introns is consistent with the consensus sequence at exon-intron junctions where an intron begins with a GT and ends with an AG dinucleotide (Breathnach and Chambon, 1981). Sequences at both ends of the introns are complementary to the 5′-terminal sequences of the U1 small nuclear RNA as found for other genes (Lerner *et al.*, 1980; Rogers and Wall, 1980). The arrow head points to the start site of transcription; the underlined regions show the Goldberg-Hogness sequence (Goldberg, 1979) and the exon-intron junctions. A CAAT sequence (Benoist *et al.*, 1980), typically 70–80 nucleotides upstream of the transcription start site and found in many but not all eukaryotic genes, could not be identified. A polyadenylation signal AAAATAAA (Proudfoot and Brownlee, 1976) is found 38 nucleotides downstream of the stop codon. The rat AVP-Np gene is 1.85 kb long and comprises three exons, A, B and C of, respectively, 163 bp, 202 bp and 210 bp.

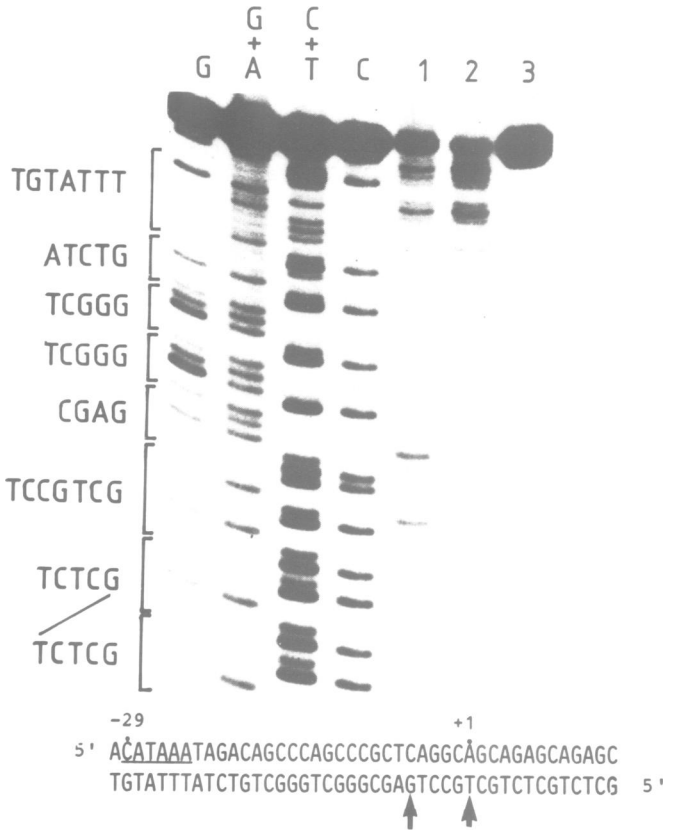

Fig. 3. Localization of the 5′ terminus of rat prepro-AVP-Np mRNA by S1 nuclease mapping. **Lanes G, G+A, C+T, C:** DNA sequence of the *Sau*961-*Pst*I fragment (sense strand; see Materials and methods). **Lane 1:** S1 nuclease-resistant products generated by digestion of hybrids between rat hypothalamic poly(A)$^+$ RNA and the genomic *Sau*961-*Pst*I fragment. Lane **2:** as in **lane 1**, except that rat hypothalamic poly(A)$^+$ RNA was substituted by 2 μg of *Escherichia coli* tRNA. **Lane 3:** the *Sau*961-*Pst*I fragment undigested. Below the figure is shown the DNA sequence of part of the *Sau*961-*Pst*I fragment. Arrows indicate the positions of the two major S1- resistant products at position −5 and +1. The Goldberg-Hogness box CATAAAT is underlined.

89 of the prepro-AVP-Np precursor. This is very close to where the highly conserved region of the rat neurophysins (Chauvet *et al.*, 1981) ends (Ser 87 of the precursor). The sequence encoding the variable C-terminal part of neurophysin and the glycoprotein as well as the 3′ non-coding region are uninterrupted. As indicated by nuclease S1 mapping (Weaver and Weissmann, 1979) there are no introns in the 5′-untranslated region nor in the sequence encoding the signal peptide (see Figure 3).

The structure of the 5′-untranslated region of the bovine cDNA shows strong homology (74%) with the rat genomic DNA upstream of the first ATG codon, the presumptive initiation site of translation. To determine the start site of transcription, poly(A)$^+$ RNA from rat hypothalami was hybridized to a 5′ fragment of the pVλ*Pvu*II-1 subclone (Figure 1) and digested with nuclease S1. A protected fragment of 88 bp (Figure 3) suggested transcription to start at position +1 (Figure 2), an A residue, as in other eukaryotic mRNAs (Breathnach and Chambon, 1981). Support for this assignment is provided by a modified 'Goldberg-Hogness' sequence, CATAAAT (Goldberg, 1979) located 29 nucleotides upstream of the presumptive transcription start site. A second protected fragment of 93 bp (Figure 3) may point to a microheterogeneity of the mRNA (Grez *et al.*, 1981), or to the formation of secondary structures not completely accessible to nuclease S1 digestion (Weaver and Weissmann, 1979).

The sequence of the rat AVP-Np gene implies a length of 575 nucleotides for its mRNA and 168 amino acids for the precursor (mol. wt. 17 826). This is in line with blot hybridization analysis of rat poly(A)$^+$ RNA (700 bases including a poly(A) tail) and cell-free translation studies (Schmale and Richter, 1981; Swann *et al.*, 1982). Dot blot analysis of RNA from the supraoptic nuclei of homozygous Brattleboro and normal rats using a specific AVP-Np probe (*Nar*I-*Hind*III) indicates that the AVP-Np mRNA is drastically reduced if not absent in the mutant strain (data not shown) agreeing with results of cell-free translation studies (Richter, 1983). Since restriction mapping of genomic DNA

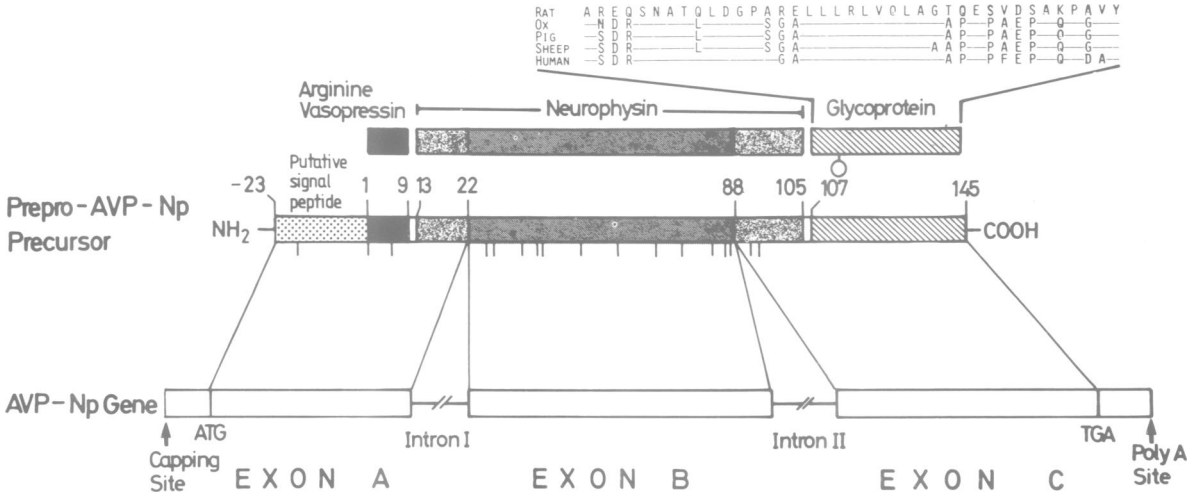

Fig. 4. Schematic representation of the structural organization of the AVP gene and the deduced protein precursor. The insertion represents the glycoprotein sequences from different sources. The rat glycoprotein sequence shows significant amino acid replacements compared with that from sheep, ox, pig and man (Seidah *et al.*, 1981; Smyth and Massey, 1979; Holwerda, 1972). The glycosylation site Asn-Ala-Thr and the leucine-rich central parts are well conserved which may be indicative of a defined function, possibly as processing signals (Smyth and Massey, 1979), preserved during evolution. Dotted bar, putative signal peptide; black bar, AVP; shaded bar, neurophysin (dark shading shows the conserved region); hatched bar, glycoprotein. The open bars in the prepro-AVP-Np precursor present the amino acids involved in post-translational processing. The positions of significant amino acid residues are numbered; cysteine residues are marked by vertical lines below the protein precursor; ♁, position of the glycosylation site. The sites of capping, translational initiation and termination and the poly(A) site are indicated.

confirms the existence of an AVP-Np gene, analysis of the genomic DNA from Brattleboro rats should give further insights into this genetic defect.

Discussion

Examination of the AVP-Np gene structure (Figure 4) identifies the principal functional domains – AVP, Np, glycoprotein – on three distinct exons similar to other split genes (Breathnach and Chambon, 1981; Gilbert, 1978; Crick, 1979). The processing signals Gly-Lys-Arg essential for the release of a functional hormone are located together with AVP on exon A. The other signal, an arginine residue between Np and the glycoprotein, is preserved in exon C. Genes encoding other polyproteins (Herbert, 1981; Numa and Nakanishi, 1981) like the pro-enkephalins (Herbert, 1981; Kakidani *et al.*, 1982; Noda *et al.*, 1982; Gubler *et al.*, 1982; Legon *et al.*, 1982) apparently adopted a mechanism where repetitive enkephalin units together with spacer regions and proteolytic cleavage signals are encoded within a single exon. In these cases pre-existing introns separating the functional units might have been lost during evolution (Perler *et al.*, 1980; Bell *et al.*, 1980) after repetition of an ancestral gene (Noda *et al.*, 1982; Gubler *et al.*, 1982; Legon *et al.*, 1982). Except for an ancient gene duplication within the neurophysin exon (B), this mechanism does not appear to have played a significant role in the evolution of the AVP-Np gene, which better fits an exon shuffling model (Gilbert, 1978; Crick, 1979). Comparison of the rat gene sequence with the bovine cDNA indicates striking homologies not only in the protein-coding sequence but also in the 5′- and 3′-untranslated regions. This finding together with the remarkable sequence homology of 197 nucleotides encoding the conserved part of bovine NpI and NpII (Land *et al.*, 1982, 1983) support the notion that AVP and OT genes were subject to a recent gene conversion event (Baltimore, 1981). At least in the rat AVP-Np gene an alternate splicing mechanism (Amara *et al.*, 1982) appears to be unlikely since the AVP-and OT-associated neurophysins show three amino acid replacements in their conserved regions.

Materials and methods

Screening of the Charon 4A library

The rat DNA library used (provided by J. Bonner through H. Land) was constructed by inserting rat liver DNA fragments from a partial *Eco*RI digest into a bacteriophage λ Charon 4A vector (Sargent *et al.*, 1979). It was screened by *in situ* hybridization at 65°C using as a probe a ^{32}P-nick-translated cDNA encoding the bovine AVP-NpII precursor (Land *et al.*, 1982) cloned in pBR322 (pVNPII-1). Hybridizing phage clones were isolated by repeated plaque purification and DNA was prepared from phage grown in liquid culture (Blattner *et al.*, 1977).

Restriction mapping, subcloning and sequence analysis

*Eco*RI restriction digests of DNA from selected positive plaques indicated two types of clones both sharing a fragment of 8.2 kb that reacted with the radioactive probe. The partial restriction map corresponding to the genomic DNA inserted in λ Charon 4A-VN 3 was deduced from analysis of DNA samples singly or multiply digested with restriction endonuclease. The relative positions of the exons were determined by blot hybridization with the ^{32}P-labeled cDNA probe (Land *et al.*, 1982). The exact location and the structure of the gene was determined by DNA sequencing (Maxam and Gilbert, 1980), which also confirmed the position of restriction cleavage sites. The 13.8-kb DNA insert of Charon 4A-VN 3 contained four *Eco*RI fragments of 8.2, 2.2, 2.19 and 1.3 kb, of which the largest hybridized with the nick-translated cDNA probe. Digestion with endonuclease *Hind*III gave rise to a 3.8-kb fragment, which was subcloned in the plasmid vector pUC 8 (Vieira and Messing, 1982) to give pVλ*Hind*III. Digestion of pVλ*Hind*III with endonuclease *Pvu*II gave rise to two strongly hybridizing fragments of 540 and 560 bp. Both *Pvu*II-fragments were subcloned in pUC 8 yielding the clones pVλ*Pvu*II-1 containing sequences from the 5′ and pVλ*Pvu*II-2 from the 3′ end of the gene and used together with pVλ*Hind*III for sequence analysis as outlined in Figure 1. Restriction mapping of rat liver DNA probed with a specific VNp-DNA (*Nar*/*Hind*III fragment) confirmed the fragment sizes identified with *Hind*III and *Eco*RI of the cloned rat genomic DNA library; there was no evidence for multiple copies of the AVP-Np gene in the rat.

S1 nuclease mapping

The genomic subclone pVλ*Pvu*II-1 was digested with *Sau*961, labeled at the 5′ ends with ^{32}P and recut with *Pst*I. The isolated 116-bp *Sau*961-*Pst*I fragment, uniquely labeled at the 5′ end of the sense strand, encodes the N-terminal 18 amino acids of the AVP-Np precursor and extends as far as the CATAAAT box in the 5′ direction. The probe (~0.3 pmol, sp. act. 1.7 x 10^5 c.p.m./pmol 5′ end) was denatured, hybridized to 2 μg of rat hypothalamic poly(A)$^+$ RNA (Richter *et al.*, 1980) in 80% formamide at 45°C for 15 h and digested with S1 nuclease (Weaver and Weissmann, 1979). A sample of the probe was subjected to sequence reactions (Maxam and Gilbert, 1980) and used as a marker. The products of S1 digestion were electrophoresed on a 7 M urea/10% polyacrylamide gel.

Acknowledgements

We thank Dr. H. Land for his help in the initial screening step of the rat DNA library; Drs. R. Ivell and H. Arnold for critical discussions during preparation of the manuscript; Dr. H. Bloecker for help in computer analysis; and Deutsche Forschungsgemeinschaft for financial support.

References

Acher,R. (1979) *Angew. Chem.*, **91**, 905-919.
Amara,S.G., Jonas,V., Rosenfeld,M.G., Ong,E.S. and Evans,R.M. (1982) *Nature*, **298**, 240-244.
Baltimore,D. (1981) *Cell*, **24**, 592-594.
Bell,G.I., Pictet,R.L., Rutter,W.J., Cordell,B., Tischer,E. and Goodman, H.M. (1980) *Nature*, **284**, 26-32.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Blattner,F.R., Williams,B.G., Blechl,A.E., Denniston-Tompson,K., Faber, H.E., Furlong,L.-A., Grunwald,D.J., Kiefer,D.O., Moore,D.D., Schumm, J.W., Sheldon,E.L. and Smithies,O. (1977) *Science (Wash.)*, **196**, 161-169.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Breslow,E. (1979) *Annu. Rev. Biochem.*, **48**, 251-274.
Brownstein,M.J., Russell,J.T. and Gainer,H. (1980) *Science (Wash.)*, **207**, 373-378.
Chauvet,M.T., Chauvet,J. and Acher,R. (1981) *Biochem. Biophys. Res. Commun.*, **103**, 595-603.
Cohen,P., Nicolas,P. and Camier,M. (1979) in *Current Topics in Cellular Regulation*, published by Academic Press, NY, **15**, 263-318.
Crick,F. (1979) *Science (Wash.)*, **204**, 264-271.
Gilbert,W. (1978) *Nature*, **271**, 501.
Goldberg,M.L. (1979) Thesis, Stanford University.
Grez,M., Land,H., Giesecke,K. and Schütz,G. (1981) *Cell*, **24**, 743-752.
Gubler,U., Seeburg,P., Hoffmann,B.J., Gage,L.P. and Udenfriend,S. (1982) *Nature*, **295**, 206-208.
Herbert,E. (1981) *Trends Biochem. Sci.*, **6**, 184-188.
Holwerda,D.A. (1972) *Eur. J. Biochem.*, **28**, 340-346.
Kakidani,H., Furutani,Y., Takahashi,H., Noda,M., Morimoto,Y., Horose, T., Asai,M., Inayama,S., Nakanishi,S. and Numa,S. (1982) *Nature*, **298**, 245-249.
Land,H., Grez,M., Ruppert,S., Schmale,H., Rehbein,M., Richter,D. and Schütz,G. (1983) *Nature*, **302**, 342-344.
Land,H., Schütz,G., Schmale,H. and Richter,D. (1982) *Nature*, **295**, 299-303.
Legon,S., Glover,D.M., Hughes,J., Lowry,P.J., Rigby,P.W.J. and Watson, C.J. (1982) *Nucleic Acids Res.*, **10**, 7905-7918.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L. and Steitz,J.A. (1980) *Nature*, **283**, 220-224.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Noda,M., Teranishi,Y., Takahashi,M., Toyosato,M., Notake,M., Nakanishi,S. and Numa,S. (1982) *Nature*, **297**, 431-434.
North,W.G., LaRochelle,F.T., Morris,J.F., Sokol,H.W. and Valtin,H. (1978) in Lederis,K. and Veale,W.L. (eds.), *Current Studies of Hypothalamic Function*, Karger, Basel, **1**, 62-76.
Numa,S. and Nakanishi,S. (1981) *Trends Biochem. Sci.*, **6**, 274-277.
Perler,F., Efstratiadis,A., Lomedico,P., Gilbert,W., Kolodner,R. and Dodgson,J. (1980) *Cell*, **20**, 555-566.

Pickering,B.T. and Swann,R.W. (1983) in Contin,M. (ed.), *The Secretory Process*, Karger, Basel, pp. 1-29.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Richter,D. (1983) in Austin,L. and Jeffrey,P.L. (eds.), *Progress in Molecular Aspects of Neurological Disorders*, Academic Press, Sydney, in press.
Richter,D., Schmale,H., Ivell,R. and Schmidt,C. (1980) in Koch,G. and Richter,D. (eds.), *Biosynthesis, Modification and Processing of Cellular and Viral Polyproteins*, Academic Press, NY, pp. 43-66.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1877-1879.
Sargent,T.D., Wu,J.-R., Sala-Trepat,J.M., Wallace,R.B., Reyes,A.A. and Bonner,J. (1979) *Proc. Natl. Acad. Sci. USA*, **78**, 243-246.
Schmale,H. and Richter,D. (1981) *Neuropeptides*, **2**, 151-156.
Seidah,N.G., Benjannet,S. and Chrétien,M. (1981) *Biochem. Biophys. Res. Commun.*, **100**, 901-907.
Smyth,D.G. and Massey,D.E. (1979) *Biochem. Biophys. Res. Commun.*, **87**, 1006-1010.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Swann,R.W., Gonzales,C.B., Birkett,S.D. and Pickering,B.T. (1982) *Biochem. J.*, **208**, 339-349.
Valtin,H., Stewart,J. and Sokol,H.W. (1974) in *Handbook of Physiology*, American Physiological Society, Washington, DC, **7**, 131-171.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.