The EMBO Journal Vol.2 No.5 pp.775–779, 1983

Isolation and sequence analysis of the human corticotropin-releasing factor precursor gene

Shigeki Shibahara, Yuuki Morimoto, Yasuji Furutani, Mitsue Notake, Hideo Takahashi, Shin Shimizu, Saburo Horikawa and Shosaku Numa*

Department of Medical Chemistry, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Communicated by D.Oesterhelt
Received on 23 February 1983

A human genomic DNA segment containing the gene for the corticotropin-releasing factor precursor has been isolated by screening a gene library with an ovine cDNA probe. The cloned DNA segment has been subjected to restriction endonuclease mapping and nucleotide sequence analysis. Comparison of the nucleotide sequence of the gene with that of the ovine cDNA indicates that an intron of 800 bp is inserted in the segment encoding the 5′-untranslated region of the mRNA. The segment corresponding to the protein-coding and the 3′-untranslated region of the mRNA is uninterrupted. The mRNA and amino acid sequences of the human corticotropin-releasing factor precursor have been deduced from the corresponding gene sequence. The deduced amino acid sequence of human corticotropin-releasing factor exhibits seven amino acid substitutions in comparison with the ovine counterpart.

Key words: corticotropin-releasing factor precursor gene/human corticotropin-releasing factor/intron/nucleotide sequence

Introduction

Corticotropin-releasing factor (CRF), which was found to exist in the hypothalamus (Guillemin and Rosenberg, 1955; Saffran and Schally, 1955), mediates the neural control of the pituitary-adrenocortical system and plays a critical role in the endocrine response to stress. Vale's group (Vale *et al.*, 1981; Spiess *et al.*, 1981) has isolated and sequenced a 41 amino acid ovine hypothalamic peptide with corticotropin-releasing activity, and this peptide is believed to represent the major physiological CRF (for reviews, see Fink, 1981; Yasuda *et al.*, 1982). Recently, we have elucidated the primary structure of the biosynthetic precurser of ovine CRF (hereafter referred to as prepro-CRF) by determining the nucleotide sequence of cloned DNA complementary to the mRNA encoding the precursor (Furutani *et al.*, 1983). To understand the molecular mechanism responsible for the hypothalamic-pituitary-adrenocortical response to stress, it is essential to isolate and characterize the gene encoding prepro-CRF and to study the regulation of expression of this gene. We have now cloned a human genomic DNA segment containing the prepro-CRF gene and have analysed the structural organization of the gene by restriction endonuclease mapping and DNA sequencing. The mRNA and amino acid sequences of human prepro-CRF have been deduced from the corresponding gene sequence.

*To whom reprint requests should be sent.

© IRL Press Limited, Oxford, England.

Results and Discussion

Cloning of the human prepro-CRF gene

A human genomic DNA library, kindly provided by T.Maniatis (Lawn *et al.*, 1978), was screened for phage carrying prepro-CRF gene sequences by hybridization *in situ* with an ovine cDNA probe. From ~5 x 10^5 plaques, three hybridization-positive clones were isolated. One of these clones (first type) carried a DNA insert containing three *Eco*RI fragments with approximate lengths of 1.2, 1.5 and 11.0 kb, and only the 11.0-kb fragment hybridized with the cDNA probe. The remaining two clones (second type) contained two *Eco*RI fragments with approximate lengths of 6.1 kb and 8.0 kb, and only the 8.0-kb fragment was hybridization-positive. Restriction endonuclease mapping and blot hybridization analysis with cDNA probes indicated that the 8.0-kb *Eco*RI fragment in the clones of the second type corresponds to the upstream 8.0-kb portion of the 11.0-kb *Eco*RI fragment in the clone of the first type. Figure 1A shows a restriction map thus constructed for this 11.0-kb *Eco*RI fragment. The 8.0-kb *Eco*RI fragment and the hybridization-positive 3.8-kb *Hind*III fragment derived from the 11.0-kb *Eco*RI fragment were subcloned in plasmid pBR322 for further analysis.

Blot hybridization analysis of human placental cellular DNA exhibited hybridization-positive fragments anticipated from the restriction map of the cloned genomic DNA (Figure 2). No additional distinct hybridization-positive bands were detected even when hybridization was carried out at 50°C. These results indicate that the cloned DNA segment containing the prepro-CRF gene retains the sequence organization found in the cellular DNA. They further suggest that there are no other closely related genes or pseudogenes in the human genome.

Structural analysis of the human prepro-CRF gene

The cloned human genomic DNA segment containing exons was subjected to detailed restriction endonuclease mapping and to nucleotide sequence analysis, which was carried out by the procedure of Maxam and Gilbert (1980) according to the strategy indicated (Figure 1B). Comparison of the human genomic DNA sequence (Figure 3) with the ovine cDNA sequence (Furutani *et al.*, 1983) has enabled us to locate an intron of 800 bp in the segment encoding the 5′-untranslated region of the mRNA 15 bp upstream of the translational initiation site. The intron begins with a GT and ends with an AG dinucleotide, following the general rule for exon-intron boundaries (Breathnach *et al.*, 1978). Furthermore, the sequences found at the boundaries are consistent with the splice junction sequences observed for other genes (Mount, 1982). Because additional potential donor and acceptor sequences (for example, residues −712 to −704 and −250 to −242 as donors; residues −350 to −335 and −205 to −190 as acceptors) are found in the intron, a possibility of alternative splicing (Early *et al.*, 1980; DeNoto *et al.*, 1981; Noda *et al.*, 1982; Amara *et al.*, 1982; Crabtree and Kant, 1982) cannot be excluded. The segment corresponding to the protein-coding and the 3′-untranslated region of the mRNA is uninterrupted.

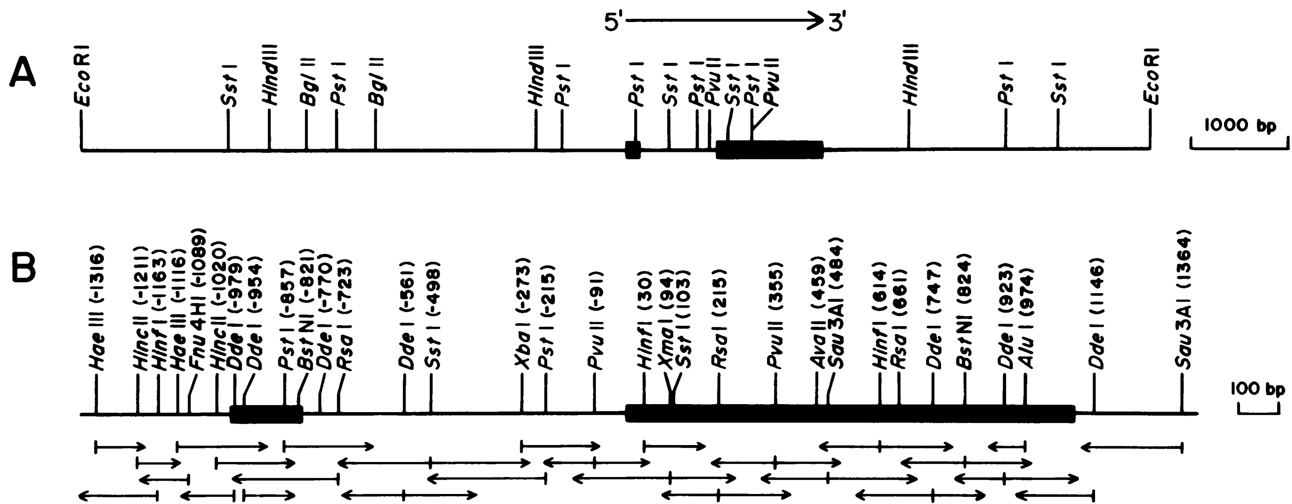

Fig. 1. Restriction mapping of the cloned human genomic DNA segment containing the prepro-CRF gene and sequencing strategy. Restriction maps with scales on the right side are given for the 11.0-kb *Eco*RI fragment **(A)** and a portion of this fragment containing exonic sequences **(B)**. The direction of transcription is from left to right. For reference, the locations of exons are indicated by closed boxes. All existing sites for the endonucleases shown are displayed in **A**, but only the relevant restriction sites are given in **B** and they are identified by numbers indicating the 5′-terminal nucleotide generated by cleavage (for the nucleotide numbers, see Figure 3). The horizontal arrows beneath the restriction map in **B** indicate the direction and extent of sequence determinations; the sites of 5′ end-labelling are shown by short vertical lines at the end of the arrows.

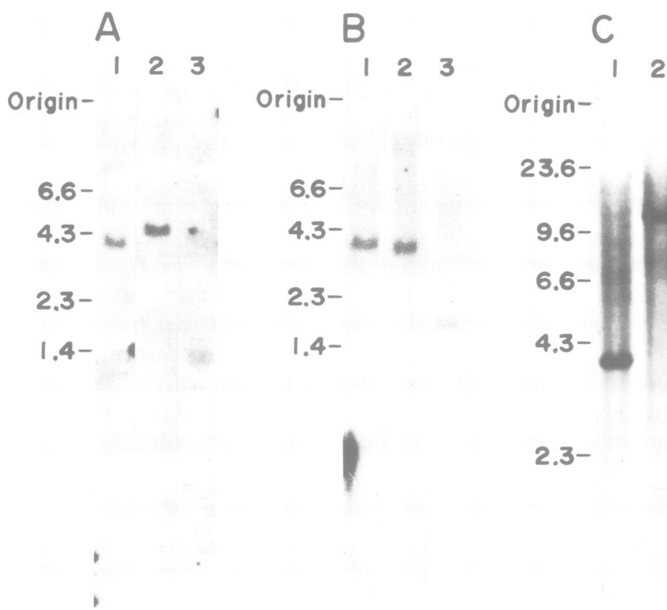

Fig. 2. Autoradiogram of blot hybridization analysis of human placental DNA. Samples of human placental DNA (15 μg each) were digested with the following endonucleases: **(A** and **B)** *Hin*dIII **(lane 1)**, *Sst*I **(lane 2)** or *Hin*dIII plus *Sst*I **(lane 3)**; **(C)** *Hin*dIII **(lane 1)** or *Eco*RI **(lane 2)**. The digestion products were electrophoresed on a 1% **(A** and **B)** or 0.7% **(C)** agarose gel, transferred to a nitrocellulose filter and hybridized at 60°C **(A** and **B)** or 50°C **(C)** to the following cloned human genomic DNA probes (see Figure 1B): **(A)** the *Hinf*I(−1163)-*Pst*I(−857) fragment containing 128 bp of the upstream exon; **(B** and **C)** the *Hinf*I(30)-*Hinf*I(614) fragment containing 584 bp of the downstream exon. The size markers, the lengths of which are given in kb, were a mixture of λcI857S7 DNA (Philippsen *et al.*, 1978) digested with *Hin*dIII and pBR322 DNA (Sutcliffe, 1978) digested with *Pst*I and *Hinf*I. On **lanes 2** and **3** in **B**, the 600-bp *Sst*I fragment was not detectable, probably because only 73 bp of this fragment are homologous with the probe used.

Attempts to identify the 5′ terminus of prepro-CRF mRNA on the cloned DNA by the S1 nuclease mapping procedure or by a primer elongation experiment were unsuccessful, apparently because of the very low content of this mRNA in hypothalamic poly(A) RNA (Furutani *et al.*, 1983). We tentatively assign the capping site to the A residue at position −985, which is located 55 bp upstream of the site corresponding to the 5′ end of the cloned ovine cDNA (Furutani *et al.*, 1983), because a putative TATA box (for review, see Breathnach and Chambon, 1981) and a putative CAAT box (Benoist *et al.*, 1980) are found 23 bp and 58 bp upstream of the putative capping site, respectively, and because eukaryotic mRNAs generally start with an A residue (for review, see Breathnach and Chambon, 1981). The translational initiation site is assigned to the AUG codon at positions 1−3 by comparison with the ovine cDNA sequence (Furutani *et al.*, 1983). This assignment is corroborated by the finding that this site represents the first AUG triplet that appears downstream of the nonsense codon UAA (residues −911 to −909) found in-frame. The 3′-untranslated region of human prepro-CRF mRNA contains two copies of the sequence AAUAAA (residues 1010−1015 and 1077−1082), which are probably involved in poly(A) addition after transcription (Proudfoot and Brownlee, 1976). It is possible that human prepro-CRF mRNA, like the ovine counterpart (Furutani *et al.*, 1983), is polyadenylated at multiple sites. The polyadenylation sites are assigned to residues 1036 and 1094 by comparison with the ovine cDNA sequence.

Primary structure of human prepro-CRF and its comparison with ovine prepro-CRF

The primary structure of human prepro-CRF has been deduced from the corresponding gene sequence (Figure 3). Human prepro-CRF consists of 196 amino acid residues. The sequence of the amino-terminal 24 amino acid residues exhibits a feature characteristic of the signal peptide of secretory proteins (Blobel and Dobberstein, 1975), which generally contains a region rich in hydrophobic amino acids with large side chains in its central portion and terminates in a residue having a small neutral side chain (for example, Ala, Gly or Ser) (Steiner *et al.*, 1980). Therefore, a possible site for cleavage of the signal peptide of human prepro-CRF seems to be located after the alanine residue at position 24 (alternatively after the serine residue at position 27).

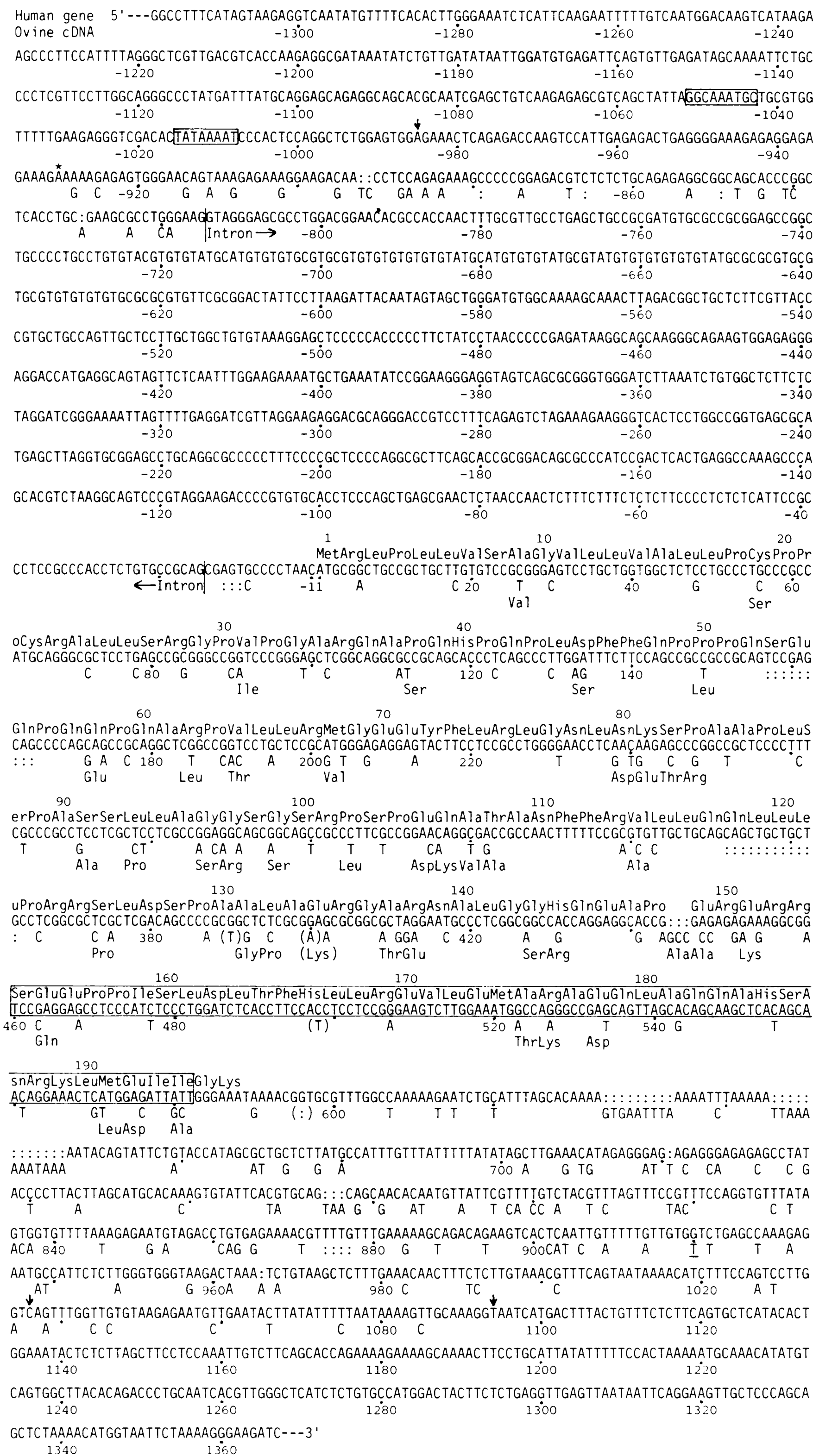

Fig. 3. Nucleotide sequence of the human prepro-CRF gene and its comparison with that of ovine cDNA. The nucleotide sequence of the human message strand, together with the deduced amino acid sequence, is shown. Nucleotide residues in the human sequence are numbered in the 5′ to 3′ direction, beginning with the first residue of the ATG triplet encoding the initiative methionine, and the nucleotides on the 5′ side of residue 1 are indicated by negative numbers. Amino acid residues in the human sequence are numbered beginning with the initiative methionine. The nucleotide and amino acid differences found in the ovine sequences (Furutani *et al.*, 1983) are displayed beneath the human sequences; the absence of a nucleotide or an amino acid in the ovine sequence means that the human and ovine sequences are the same; the presence of a colon in either nucleotide sequence indicates a gap; the nucleotide or amino acid differences in parentheses mean that they apply only to a certain ovine cDNA clone(s), whereas the other clone(s) shows the same sequence as the human counterpart; at position 920, T (underlined) or C is found in ovine cDNA clones instead of G in the human sequence. The 5′ end of the cloned ovine cDNA is marked with an asterisk. The exon-intron junctions, which are positioned according to the GT/AG rule (Breathnach *et al.*, 1978), are indicated by vertical lines. The putative sites of capping and poly(A) addition are shown by arrows. The putative TATA box and the putative CAAT box are enclosed with boxes. The sequence of human CRF, together with the coding nucleotides, is also boxed.

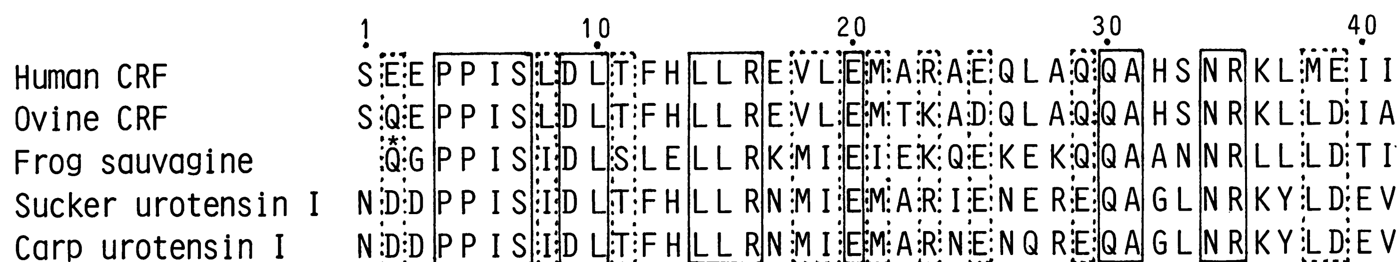

Fig. 4. Comparison of the amino acid sequences of human CRF, ovine CRF, frog sauvagine, sucker urotensin I and carp urotensin I. The sequence data have been taken from the following references: ovine CRF (Vale *et al.*, 1981); *Phyllomedusa sauvagei* sauvagine (Montecucchi *et al.*, 1979); *Catostomus commersoni* urotensin I (Lederis *et al.*, 1982a); *Cyprinus carpio* urotensin I (Ichikawa *et al.*, 1982). The one-letter amino acid notation is used. For comparison, the pyrrolidonecarboxylic acid residue at the amino terminus of sauvagine is regarded as glutamine (marked with an asterisk). The carboxy-terminal residues of all the peptides are amidated; the carboxy-terminal amidation of human CRF is assumed (see text). Sets of five identical residues are enclosed with solid lines, and sets of five residues considered to be favoured amino acid substitutions are enclosed with dotted lines. Favoured amino acid substitutions are defined as pairs of residues belonging to one of the following groups: S, T, P, A and G; N, D, E and Q; H, R and K; M, I, L and V; F, Y and W (Dayhoff *et al.*, 1978).

The deduced structure of human CRF, which is composed of 41 amino acids (residues 154 – 194), exhibits seven amino acid substitutions in comparison with ovine CRF (Vale *et al.*, 1981), that is, Glu for Gln (2), Ala for Thr (22), Arg for Lys (23), Glu for Asp (25), Met for Leu (38), Glu for Asp (39) and Ile for Ala (41); the numbers in parentheses indicate the amino acid number of CRF beginning with its amino terminus. All these replacements represent changes among chemically similar amino acids (Dayhoff *et al.*, 1978) and are caused by single nucleotide substitutions, except for the Ile/Ala replacement at the carboxyl terminus of CRF. The sequence of human CRF in its precursor, like the ovine counterpart (Furutani *et al.*, 1983), is preceded by the paired basic residues Arg-Arg, which apparently represent the site of proteolytic processing (Steiner *et al.*, 1980), and is followed by the dipeptide Gly-Lys, which constitutes the carboxyl end of prepro-CRF. This implies that the carboxy-terminal isoleucine residue of human CRF is also amidated (Suchanek and Kreil, 1977; Nakanishi *et al.*, 1979; Amara *et al.*, 1980; Shibasaki *et al.*, 1980; Seidah *et al.*, 1981; Land *et al.*, 1982; Yoo *et al.*, 1982; Bradbury *et al.*, 1982; Furutani *et al.*, 1983).

Comparison of the amino acid sequences of human and ovine prepro-CRF shows that the CRF and the signal peptide region are more highly conserved than the remaining regions. The degree of homology is 83% for CRF (residues 154 – 194) and 92% for the signal peptide (assuming the segment of residues 1 – 24), whereas it is 73% for the region composed of residues 25 – 122 and 64% for the region composed of residues 125 – 151; the latter two regions are separated by the paired basic residues Arg-Arg which may represent a site of proteolytic processing. These less conserved regions exhibit seven amino acid additions (Ser-Glu-Gln at positions 53 – 55 and Gln-Leu-Leu-Leu at positions 118 – 121) and one amino acid deletion (Ala between positions 148 and 149) in the human sequence as compared with the ovine counterpart; for calculating the degree of homology, these gaps have been counted as one substitution regardless of their length.

Figure 4 compares the amino acid sequences of human and ovine CRF, sauvagine from frog skin (Montecucchi *et al.*, 1979) and urotensin I from sucker and carp urophysis (Lederis *et al.*, 1982a; Ichikawa *et al.*, 1982). The two nonmammalian peptides also stimulate the release of corticotropin and exhibit potent vasodilatory, hypotensive activity (Lederis *et al.*, 1982a, 1982b); the vasodilatory activity of CRF is considerably weaker than those of urotensin I and sauvagine (Lederis *et al.*, 1982b). The regions near the amino and the carboxyl termini are highly conserved among the five peptides and may be essentially involved in corticotropin-releasing activity. Computer analysis of the amino acid sequences for secondary structure (Chou and Fasman, 1978) shows that the segments extending from position 7 to position 31 of these peptides share a predicted α-helical structure. Secondary structure predictions of ovine CRF and sauvagine have been reported (Montecucchi and Gozzini, 1982).

Materials and methods

Screening of the human genomic DNA library

The gene library used was a collection of recombinant phage that carried human fetal liver DNA fragments generated by partial digestion with *Hae*III and *Alu*I and joined to the λ Charon 4A arms with *Eco*RI linkers (Lawn *et al.*, 1978). Phage were screened by hybridization *in situ* according to the procedure of Benton and Davis (1977) modified by Woo *et al.* (1978). The hybridization probe used was the *Rsa*I fragment of ~650 bp containing nucleotide residues 110 – 724 of ovine prepro-CRF cDNA, which was excised from clone pCRF31 (Furutani *et al.*, 1983) and was labelled with [α-^{32}P]dCTP by nick-translation (Weinstook *et al.*, 1978); one of the *Rsa*I sites was located on the vector DNA (Okayama and Berg, 1982) 22 bp upstream of the *Pst*I site flanking the cDNA insert. Hybridization-positive phage clones were isolated by repeated plaque purification.

Restriction mapping, blot hybridization and nucleotide sequence analysis

Restriction endonucleases were purchased from Takara Shuzo Co., Bethesda Research Laboratories and New England BioLabs. Reactions were carried out under the conditions recommended by the suppliers. Separation of the restriction fragments was carried out by electrophoresis on agarose gel (Polsky *et al.*, 1978) or on polyacrylamide gel (Nakanishi *et al.*, 1979). Blot hybridization analysis of human placental DNA, isolated by the procedure of Polsky *et al.* (1978), and of cloned genomic DNA was conducted according to the methods described by Southern (1975) and Smith and Summers (1980), respectively. The ovine cDNA probes used were the *Rsa*I fragment described for screening phage (see above) and the *Alu*I-*Ava*I fragment containing nucleotide residues – 127 to 93 of prepro-CRF cDNA derived from pCRF43 (Furutani *et al.*, 1983); the *Alu*I site is located on the vector DNA (Sutcliffe, 1978) 48 bp upstream of the *Pst*I site flanking the cDNA insert. For the human genomic DNA probes used, see the legend to Figure 2. The hybridization probes were labelled by nick-translation (see above). 5′ End-labelling of restriction fragments and DNA sequencing were carried out by the procedure of Maxam and Gilbert (1980). Subcloning of genomic DNA fragments in plasmid pBR322 was performed as described by Frischauf *et al.* (1980).

Acknowledgements

We thank Dr.T.Maniatis for the human genomic DNA library and Dr.T.Miyata for computer analysis. This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

References

Amara,S.G., David,D.N., Rosenfeld,M.G., Roos,B.A. and Evans,R.M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4444-4448.
Amara,S.G., Jonas,V., Rosenfeld,M.G., Ong,E.S. and Evans,R.M. (1982) *Nature*, **298**, 240-244.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Blobel,G. and Dobberstein,B. (1975) *J. Cell Biol.*, **67**, 852-862.
Bradbury,A.F., Finnie,M.D.A. and Smyth,D.G. (1982) *Nature*, **298**, 686-688.
Breathnach,R., Benoist,C., O'Hare,K., Gannon,F. and Chambon,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4853-4857.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Chou,P.Y. and Fasman,G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251-276.
Crabtree,G.R. and Kant,J.A. (1982) *Cell*, **31**, 159-166.
Dayhoff,M.O., Schwartz,R.M. and Orcutt,B.C. (1978) in *Atlas of Protein Sequence and Structure, 5, Suppl. 3,* National Biomedical Research Foundation, Silver Spring, MD, pp. 345-352.
DeNoto,F.M., Moore,D.D. and Goodman,H.M. (1981) *Nucleic Acids Res.*, **9**, 3719-3730.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **20**, 313-319.
Fink,G. (1981) *Nature*, **294**, 511-512.
Frischauf,A.M., Garoff,H. and Lehrach,H. (1980) *Nucleic Acids Res.*, **8**, 5541-5549.
Furutani,Y., Morimoto,Y., Shibahara,S., Noda,M., Takahashi,H., Hirose, T., Asai,M., Inayama,S., Hayashida,H., Miyata,T. and Numa,S. (1983) *Nature*, **301**, 537-540.
Guillemin,R. and Rosenberg,B. (1955) *Endocrinology*, **57**, 599-607.
Ichikawa,T., McMaster,D., Lederis,K. and Kobayashi,H. (1982) *Peptides*, **3**, 859-867.
Land,H., Schütz,G., Schmale,H. and Richter,D. (1982) *Nature*, **295**, 299-303.
Lawn,R.M., Fritsch,E.F., Parker,R.C., Blake,G. and Maniatis,T. (1978) Cell, **15**, 1157-1174.
Lederis,K., Letter,A., McMaster,D., Moore,G. and Schlesinger,D. (1982a) *Science (Wash.)*, **218**, 162-164.
Lederis,K., Vale,W., Rivier,J., MacCannell,K.L., McMaster,D., Kobayashi, Y., Suess,U. and Lawrence,J. (1982b) *Proc. West. Pharmacol. Soc.*, **25**, 223-227.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Montecucchi,P.C. and Gozzini,L. (1982) *Int. J. Peptide Protein Res.*, **20**, 139-143.
Montecucchi,P.C., Henschen,A. and Erspamer,V. (1979) *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 1178.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459-472.
Nakanishi,S., Inoue,A., Kita,T., Nakamura,M., Chang,A.C.Y., Cohen,S.N. and Numa,S. (1979) *Nature*, **278**, 423-427.
Noda,M., Teranishi,Y., Takahashi,H., Toyosato,M., Notake,M., Nakanishi, S. and Numa,S. (1982) *Nature*, **297**, 431-434.
Okayama,H. and Berg,P. (1982) *Mol. Cell. Biol.*, **2**, 161-170.
Philippsen,P., Kramer,R.A. and Davis,R.W. (1978) *J. Mol. Biol.*, **123**, 371-386.
Polsky,F., Edgell,M.H., Seidman,J.G. and Leder,P. (1978) *Anal. Biochem.*, **87**, 397-410.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Saffran,M. and Schally,A.V. (1955) *Can. J. Biochem. Physiol.*, **33**, 408-415.
Seidah,N.G., Rochemont,J., Hamelin,J., Benjannet,S. and Chrétien,M. (1981) *Biochem. Biophys. Res. Commun.*, **102**, 710-716.
Shibasaki,T., Ling,N. and Guillemin,R. (1980) *Biochem. Biophys. Res. Commun.*, **96**, 1393-1399.
Smith,G.E. and Summers,M.D. (1980) *Anal. Biochem.*, **109**, 123-129.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Spiess,J., Rivier,J., Rivier,C. and Vale,W. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6517-6521.
Steiner,D.F., Quinn,P.S., Chan,S.J., Marsh,J. and Tager,H.S. (1980) *Ann. N.Y. Acad. Sci.*, **343**, 1-16.
Suchanek,G. and Kreil,G. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 975-978.
Sutcliffe,J.G. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77-90.
Vale,W., Spiess,J., Rivier,C. and Rivier,J. (1981) *Science (Wash.)*, **213**, 1394-1397.
Weinstock,R., Sweet,R., Weiss,M., Cedar,H. and Axel,R. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1299-1303.
Woo,S.L.C., Dugaiczyk,A., Tsai,M.-J., Lai,E.C., Catterall,J.F. and O'Malley,B.W. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3688-3692.
Yasuda,N., Greer,M.A. and Aizawa,T. (1982) *Endocrinol. Rev.*, **3**, 123-140.
Yoo,O.J., Powell,C.T. and Agarwal,K.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1049-1053.