Cloning and sequence analysis of cDNA for rat corticotropin-releasing factor precursor

Hisato Jingami*†, Noboru Mizuno, Hideo Takahashi*†, Shigeki Shibahara*, Yasuji Furutani*, Hiroo Imura† and Shosaku Numa*°

*Departments of *Medical Chemistry and Molecular Genetics, †Medicine* and *Anatomy, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606, Japan*

Received 2 September 1985

DNA complementary to the rat hypothalamic mRNA coding for the corticotropin-releasing factor precursor (prepro-CRF) has been cloned by screening a cDNA library with a human genomic DNA probe. Nucleotide sequence analysis of the cloned cDNA has revealed that rat prepro-CRF consists of 187 amino acid residues including a putative signal peptide. The CRF and putative signal peptide regions are more highly conserved among rat, human and ovine prepro-CRF than is the cryptic portion.

Corticotropin-releasing factor *Hormone precursor* *cDNA cloning* *Nucleotide sequence*

1. INTRODUCTION

The primary structures of the biosynthetic precursors of ovine and human CRF have previously been elucidated by cloning and sequencing the cDNA and the genomic DNA, respectively [1,2]. The deduced structure of human CRF exhibits 7 amino acid substitutions in comparison with ovine CRF [3], but is identical with the structure of rat CRF [4]. In view of the fact that the rat has been used most frequently in studies on the control of the hypothalamo-pituitary-adrenocortical system, we have now cloned and sequenced cDNA for the rat prepro-CRF.

2. MATERIALS AND METHODS

Total RNA was extracted from rat hypothalami, and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography as in [1]. A cDNA library was constructed with 4 μg of this poly(A)$^+$ RNA and 1.5 μg vector-primer DNA according to [5]. *Escherichia coli* χ1776 or HB101 was used for transformation as in [1]. Ampicillin-resistant transformants were screened as in [6] by hybridization at 55°C with the 446 bp *Rsa*I fragment comprising nucleotide residues 215–660 of the human prepro-CRF gene [2]; the probe was labelled with [α-^{32}P]dCTP by nick-translation [7]. DNA sequencing was carried out by the procedure in [8]. Reagents were obtained as in [9].

Abbreviations: CRF, corticotropin-releasing factor; prepro-CRF, corticotropin-releasing factor precursor

° To whom correspondence should be addressed

Dedicated to Professor Karl Decker on the occasion of his 60th birthday

3. RESULTS AND DISCUSSION

A cDNA library derived from rat hypothalamic poly(A)$^+$ RNA was screened with a human genomic DNA probe. From about 1.4×10^6 transformants, 5 hybridization-positive clones were isolated. One of them (clone prCRF87), which apparently carried the largest cDNA insert, was subjected to restriction mapping and nucleotide sequence analysis.

Fig.1 shows the nucleotide sequence of the mRNA coding for rat prepro-CRF, deduced from

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

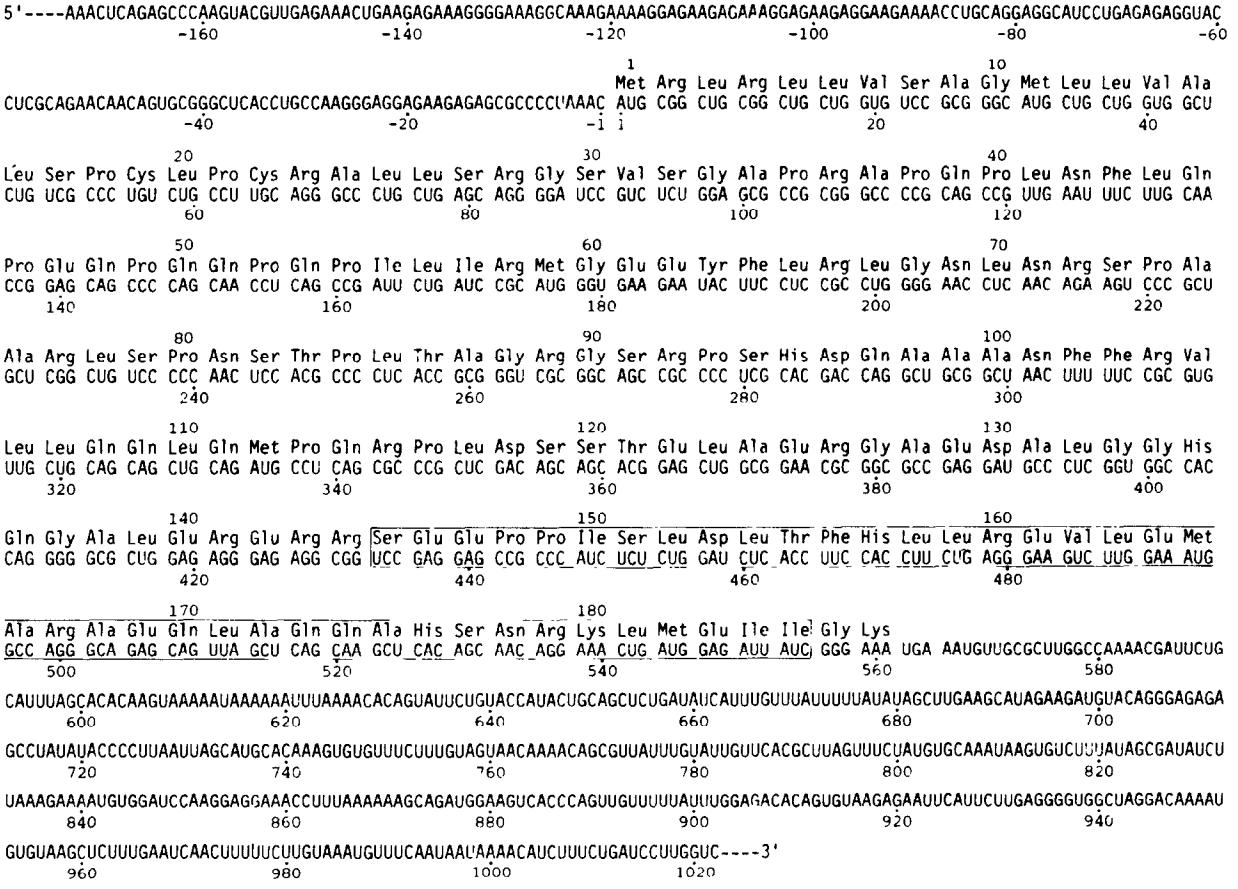

Fig.1. Primary structure of rat prepro-CRF mRNA. Nucleotide residues are numbered in the 5'- to 3'-direction, beginning with the first residue of the AUG triplet encoding the initiating methionine, and the nucleotides on the 5'-side of residue 1 are indicated by negative numbers. The 3'-terminal sequence shown is followed by a poly(A) tract. Amino acid residues are numbered beginning with the initiating methionine. The CRF sequence, together with the coding nucleotides, is boxed.

the cDNA sequence. Because the 5'-end of the cDNA sequence corresponds to the position 2 nucleotides downstream of the putative capping site of the human prepro-CRF gene [2], it is assumed that clone prCRF87 carries an almost full-length cDNA sequence. This is also supported by the size of the mRNA (approx. 1300 nucleotides) estimated by blot hybridization analysis of rat hypothalamic poly(A)$^+$ RNA [10]. The 3'-untranslated region of the mRNA contains 2 copies of the polyadenylation signal AAUAAA [11] (residues 612–617 and 996–1001), one of which is located 22 nucleotides upstream of the poly(A) tract.

The primary structure of rat prepro-CRF was deduced by using the reading frame of the mRNA corresponding to the known amino acid sequences of the ovine and human counterparts [1,2] (fig.1). The translational initiation site assigned is corroborated by the finding that this site represents the first AUG triplet that appears downstream of a nonsense codon (UGA at positions −144 to −142) found in frame. The sequence of the 24 amino acid residues starting with the initiating methionine exhibits a feature characteristic of the signal peptide of secretory proteins [12,13]. A possible site for cleavage of the signal peptide seems to be located after the alanine residue specified by the 24th codon (alternatively after the serine residue specified by the 27th codon). A translational termination codon (UGA) is found in

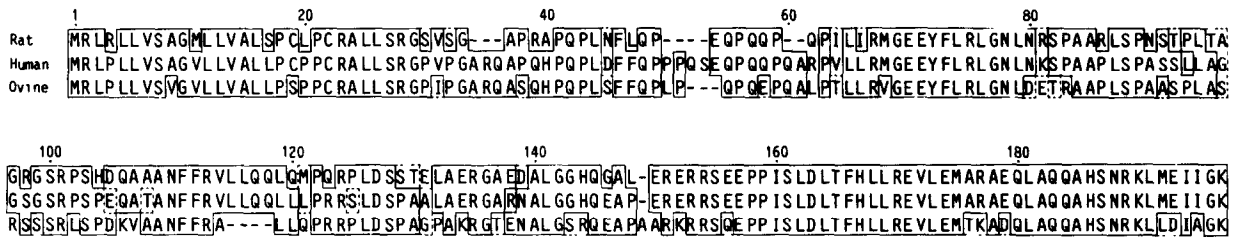

Fig.2. Alignment of the amino acid sequences of rat (top), human (middle) and ovine (bottom) prepro-CRF. The one-letter amino acid notation is used. The human and ovine sequences have been taken from [2] and [1], respectively. Sets of identical residues are enclosed with solid lines, and sets of conservative residues with dashed lines. Conservative 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 [16]. Gaps (−) have been inserted to achieve maximum homology. The positions in the aligned sequences including gaps are numbered beginning with that of the initiating methionine.

frame after the 187th codon specifying lysine. Thus, it is concluded that rat prepro-CRF consists of 187 amino acid residues including a putative signal peptide.

The deduced sequence of amino acid residues 145–185 agrees precisely with the sequence of rat CRF determined by peptide analysis [4]. The sequence of rat CRF in the precursor, like its ovine and human counterparts [1,2], is preceded by the paired basic residues Arg-Arg, which apparently represent the site of proteolytic processing [13], and is followed by the dipeptide Gly-Lys, which constitutes the carboxyl-end of the precursor. This implies that the carboxy-terminal isoleucine residue of rat CRF is also amidated (e.g., see [14,15]). The cryptic portion of ovine and human prepro-CRF contains an additional pair of basic residues (Arg-Arg) at an equivalent position (aligned positions 123 and 124 in fig.2). In rat prepro-CRF, this dibasic structure is replaced by Gln-Arg.

The alignment of the amino acid sequences of rat, human and ovine prepro-CRF is shown in fig.2. The degrees of sequence homology between the whole precursors are 80, 65 and 76% for the rat/human, rat/ovine and human/ovine pairs, respectively; gaps have been counted as one substitution regardless of their length. The CRF region (83–100%) and the putative signal peptide region (75–92%) are more highly conserved among the 3 species than is the cryptic portion corresponding to aligned positions 25–152 (55–72%), where several amino acid deletions (or insertions) occur.

ACKNOWLEDGEMENTS

We thank Drs Paul Berg and Hiroto Okayama for providing their high-efficiency cloning system, Dr Takashi Miyata for computer analysis and Dr Shigetada Nakanishi for helpful advice. 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

[1] 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.

[2] Shibahara, S., Morimoto, Y., Furutani, Y., Notake, M., Takahashi, H., Shimuzu, S., Horikawa, S. and Numa, S. (1983) EMBO J. 2, 775–779.

[3] Vale, W., Spiess, J., Rivier, C. and Rivier, J. (1981) Science 213, 1394–1397.

[4] Rivier, J., Spiess, J. and Vale, W. (1983) Proc. Natl. Acad. Sci. USA 80, 4851–4855.

[5] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.

[6] Hanahan, D. and Meselson, M. (1980) Gene 10, 63–67.

[7] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.

[8] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[9] Jingami, H., Nakanishi, S., Imura, H. and Numa, S. (1984) Eur. J. Biochem. 142, 441–447.

[10] Jingami, H., Matsukura, S., Numa, S. and Imura, H. (1985) Endocrinology 117, in press.
[11] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[12] Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67, 852–862.
[13] Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980) Ann. NY Acad. Sci. 343, 1–16.
[14] Suchanek, G. and Kreil, G. (1977) Proc. Natl. Acad. Sci. USA 74, 975–978.
[15] Bradbury, A.F., Finnie, M.D.A. and Smyth, D.G. (1982) Nature 298, 686–688.
[16] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Silver Spring, MD.