Volume 246, number 1,2, 193–196 FEB 06970 March 1989

A new species of enkephalin precursor mRNA with a distinct 5′-untranslated region in haploid germ cells

Kazuaki Yoshikawa, Kei Maruyama*, Takako Aizawa and Akihiro Yamamoto

Department of Molecular Biology, Psychiatric Research Institute of Tokyo, Kamikitazawa, Setagaya-ku, Tokyo 156 and **Department of Pharmacology, Faculty of Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan*

Received 6 February 1989

To elucidate the primary structure of preproenkephalin (A) mRNA expressed by haploid germ cells (round spermatids) in rat testis, we have screened a λgt11 cDNA library for preproenkephalin cDNA inserts. The largest cDNA insert contained a protein-coding sequence encoding 269 amino acid residues as well as 327 and 309 bases of the 5′- and 3′-untranslated regions, respectively. The protein-coding region plus 3′-untranslated region of the mRNA was over 99% homologous to that of brain preproenkephalin mRNA, whereas the 5′-untranslated region contained a distinct sequence including a partial sequence of intron A of the preproenkephalin gene [(1984) J. Biol. Chem. 259, 14301–14308; (1984) J. Biol. Chem. 259, 14309–14313]. Northern blot analysis using a 5′-end-specific probe showed that this type of preproenkephalin mRNA exists exclusively in the germ cells.

Preproenkephalin; mRNA; Untranslated region, 5′-; Intron A; Neural cell lineage; (Rat spermatid)

1. INTRODUCTION

Enkephalins (Met- and Leu-enkephalin) are widely distributed in the peripheral and central nervous systems. Recently, mRNA encoding Met-enkephalin precursor, preproenkephalin (ppEnk), has been found not only in the nervous systems but also in various non-neuronal tissues such as lymphocytes [1], heart [2], and reproductive systems [3]. In the testis, spermatogenic germ cells [4,5] and Sertoli cells [6] contain ppEnk mRNA. The size of ppEnk mRNA in the germ cells is 1.7–1.9 kb, which is larger than those found in other tissues (1.45–1.5 kb) [4,5]. This specific type of ppEnk mRNA can be translatable into a precursor because Met-enkephalin immunoreactive peptides are generated by limited proteolysis of the extracts of haploid germ cells [5]. We report here the primary structure of this germ cell-specific ppEnk mRNA, which contains a distinct 5′-untranslated sequence derived from an intervening sequence of the ppEnk gene.

Correspondence address: K. Yoshikawa, Department of Molecular Biology, Psychiatric Research Institute of Tokyo, 2-1-8, Kamikitazawa, Setagaya-ku, Tokyo 156, Japan

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no.Y07503

2. MATERIALS AND METHODS

2.1. *Preparation of haploid germ cells*

Haploid germ cells (round spermatids) were purified from testes of 125-day-old rats (Wistar strain) according to the Percoll centrifugation method in [5]. The DNA content (5.1 μg DNA/10^6 cells) of the haploid cell preparation was 53% of rat diploid fibroblasts.

2.2. *Construction of haploid germ cell cDNA library*

Total cellular RNA was purified from the germ cell preparation by the differential ethanol precipitation method [7]. Poly(A)$^+$ RNA was purified from total RNA by oligo(dT)-cellulose affinity chromatography. Oligo(dT)-primed double-stranded cDNA was synthesized from poly(A)$^+$ RNA, ligated with linkers, and cloned into the *Eco*RI site of λgt11 vector (Amersham).

2.3. *Isolation of ppEnk cDNA and sequence analysis*

Phage clones carrying ppEnk cDNA were screened by

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

hybridization with an *SmaI/SacI* fragment prepared from rat brain ppEnk cDNA (pRPE2) [8]. The largest cDNA insert prepared from one of the hybridization-positive phage clones was subcloned into the pSPT18 plasmid for DNA preparation, and appropriate restriction fragments were then prepared for nucleotide sequencing. The sequences were determined by the dideoxy chain-termination method [9] using both M13 (Toyobo) and pBluescript II (Stratagene) systems. The cDNA fragments were sequenced in both directions from *Eco*RI-cleaved ends and from three internal *Pst*I sites. The sequence data obtained were analyzed using a computer program (Genetyx, Software Development, Tokyo).

2.4. *Northern blot analysis*

Total RNA was prepared from the following tissues and cells as described previously: two brain regions (striatum and frontal cortex) of a male Wistar rat (90 days old) [10], germ cells (round spermatids) [5], cultured Sertoli cells treated with follicle-stimulating hormone (FSH) (0.5 μg/ml, 12 h) [6], and cultured C6 glioma cells [11]. Purified RNA was electrophoresed on a 6% formaldehyde/1.2% agarose gel, and blotted onto a nylon membrane (Hybond N, Amersham). RNA blots were hybridized with ^{32}P-labeled probes under the conditions in [8].

3. RESULTS

About 50000 phage clones in the haploid germ cell cDNA library were screened for ppEnk cDNA sequences. Seven clones showed positive hybridization signals, of which three were found to contain the largest inserts (about 1.5 kb), which all showed identical *Pst*I restriction maps. Thus, we prepared the cDNA insert from one of the three clones, designated λGE4, and analyzed its nucleotide sequence.

Fig.1 shows the entire nucleotide sequence of the λGE4 cDNA. The size was 1446 bases excluding 36 bases of the poly(A) tail. Overall comparison with the previously determined sequence of pRPE2 (i.e. a full-length ppEnk cDNA from rat brain library) revealed that the sequence of nucleotides −3 to 1119 of the λGE4 cDNA is more than 99% homologous to the corresponding region of pRPE2 (nucleotides −3 to 1122); there are 4 mismatches at positions −2, 33, 34 and 829, whereas three nucleotides (UGU) at positions 1020–1022 in the pRPE2 cDNA are absent from the λGE4 cDNA. In contrast, the 5′-untranslated region upstream from nucleotide −4 showed no homology to the 5′-end region of the pRPE2 cDNA (nucleotides −155 to −1). Interestingly, the transition point between nucleotides −4 and −3 corresponds precisely to the junction between intron A (i.e. the intervening sequence between exons I and II) and exon II of the rat ppEnk gene [12]. Moreover, the sequence of nucleotides −108 to −4 (the 105 bases boxed in fig.1) showed 98% homology (2 mismatches) to the partially determined 3′-terminal sequence of intron A [12]. Since the complete sequence of intron A of the ppEnk gene is unavailable at present, it is unclear whether the rest of the 5′-end region (nucleotides −327 to −109) is homologous to an undetermined sequence of intron A. On the other hand, this region showed no similarity to the previously determined sequence of the rat ppEnk gene including the 5′-region of intron A (partially determined 129 bases), 153 bases of exon I and 444 bases of the 5′-flanking region containing the promoter sequence [12]. Thus, it seems likely that the 5′-end region (nucleotides −327 to −109) of germ cell-specific ppEnk mRNA is derived from an as yet undetermined portion of intron A.

The amino acid sequence deduced from the nucleotide sequence of λGE4 cDNA is identical to that of the brain ppEnk [8] except for one amino acid substitution; the nucleotide substitution at position 34 alters Leu (pRPE2 cDNA) to Val (λGE4 cDNA) at amino acid residue 12. ppEnk encoded by the λGE4 cDNA consists of 269 amino acid residues including a signal sequence at the amino-terminal as well as biologically active opioid peptides (i.e. Met-enkephalin, Leu-enkephalin, Met-enkephalin-Arg-Gly-Leu, and Met-enkephalin-Arg-Phe), which are flanked by pairs of basic amino acids (Lys, Arg).

There are 5 sites of AUG triplet in the putative 3′-untranslated region (nucleotides −327 to −1); open frames beginning with these AUG sites are discontinued by translation termination codons localized upstream from the initiating Met codon (positions 1–3). Thus, it is unlikely that ppEnk synthesized by haploid germ cells contains an extra sequence at the amino-terminal.

Northern blot analysis using a 5′-end-specific probe showed that a single 1.7 kb species of ppEnk mRNA is present exclusively in the germ cells, but not in FSH-stimulated Sertoli cells, striatum, cerebral cortex, or C6 rat glioma cells (fig.2, upper panel). On the other hand, these regions and cells other than the germ cells contained the 1.5 kb species of ppEnk mRNA [6,10,11] as detected with a 3′-end-specific probe (fig.2, lower panel).

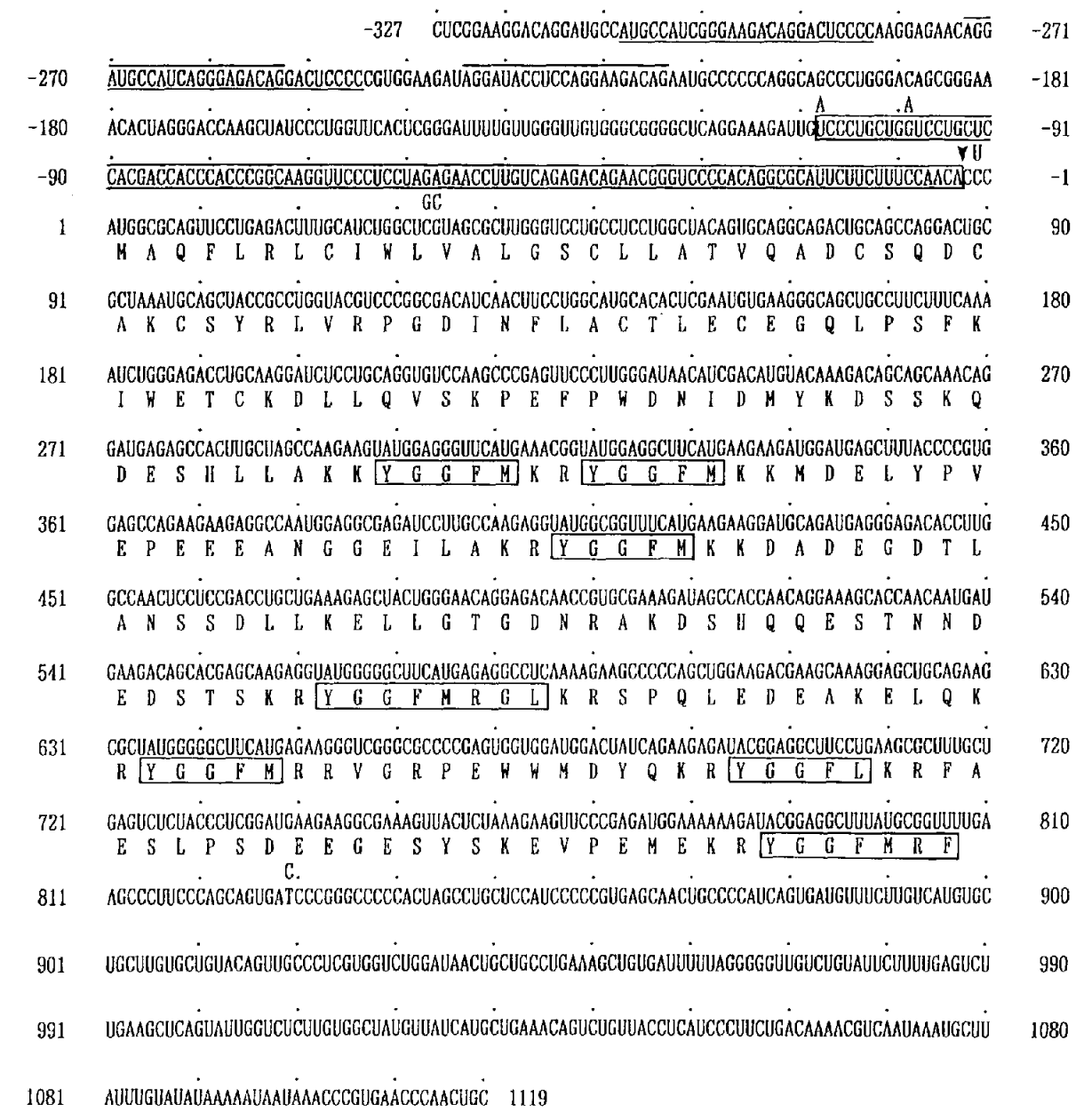

Fig.1. Primary structure of ppEnk mRNA in haploid germ cells. The nucleotide sequence of ppEnk mRNA was deduced from that of the λGE4 cDNA. Nucleotide residues are numbered in the 5′- to 3′-direction, beginning with the first residue of the AUG triplet encoding the initiating Met, which has been tentatively assigned based on the analogy to rat brain ppEnk [8]. The single-letter amino acid notation is used. The boxed amino acid sequences are of enkephalins and related opioid peptides. The arrowhead indicates the position of the intron A/exon II junction, and the boxed nucleotide sequence is highly homologous to the partially determined 3′-end sequence of intron A of rat ppEnk gene [12]. Two pairs of direct repeats of 26 bases (92% homology) and 21 bases (81% homology) in the 5′-terminal region are underlined and overlined, respectively. The nucleotides displayed above the mRNA sequence are mismatches seen in rat brain ppEnk cDNA (nucleotides −2, 33, 34 and 829) [8] and rat ppEnk gene (nucleotides −99 and −108) [12].

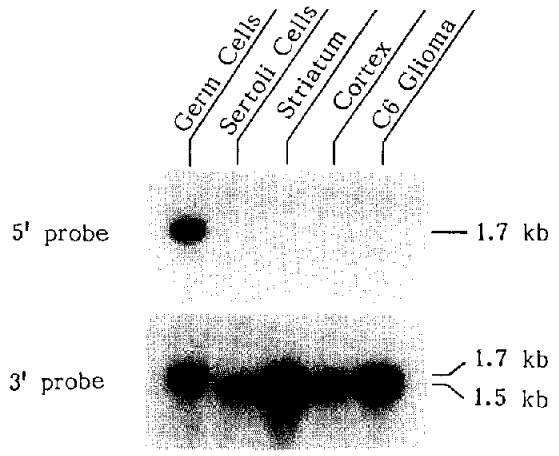

Fig.2. Northern blot analysis of ppEnk mRNA in various cells and tissues using 5'- and 3'-end-specific probes. ppEnk mRNA in total RNA (2 μg/lane) prepared from each cell or tissue was analyzed by Northern blotting. The RNA blots were hybridized with ^{32}P-labeled probes specific for either the 5'-end (5' probe, upper panel) or the 3'-end (3' probe, lower panel), which were a 307-base *Eco*RI/*Hha*I fragment (nucleotides −327 to −21 of the sequence shown in fig.1) and a 530-base *Pst*I/*Eco*RI fragment (nucleotides 626–1119 plus 36 bases poly(A) tail), respectively. Numbers to the right are the sizes of hybridized signals.

4. DISCUSSION

The present study has shown that the ppEnk mRNA expressed by haploid germ cells encodes a precursor identical to brain ppEnk except for one amino acid substitution. This substitution may be due to a strain difference because ppEnk cDNAs from Fischer rat, reported independently by us [8] and others [13], contain CTGCTA (Leu-Leu) at nucleotides 32–37, whereas ppEnk cDNA from Wistar rat (this report) and exon II of the ppEnk gene of rat (unknown strain) [12] contain CTCGTA (Leu-Val) at the corresponding positions.

The mechanism that generates the germ cell-specific ppEnk mRNA remains to be elucidated by future studies. Transcription of the ppEnk gene in the germ cells may require the conventional promoter in the 5'-flanking region of the ppEnk gene [12], but the region derived from intron A may remain unspliced due to unusual RNA splicing operating in the germ cells. Another possibility is that transcription of the ppEnk gene in the germ cells uses a distinct promoter located inside intron A.

Like mature enkephalinergic neurons, C6 glioma cells [11] and astroglial cells in primary culture [14] contain a large amount of ppEnk mRNA. This raises the possibility that the ppEnk gene is potentially expressed by immature cells during early periods of neuron-glia development. Several lines of evidence suggest a close linkage between germ cells and neural cells (review [15]). Thus, expression of the ppEnk gene by haploid germ cells may expand our knowledge concerning molecular mechanisms underlying the generation and development of cell lineage in the nervous systems.

Acknowledgements: We are grateful to Professor Yosiaki Nonomura (University of Tokyo) for continuous support of our research project. This work is supported in part by a Grant-in-Aid for General Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Zurawski, G., Benedik, M., Kamb, B.J., Abrams, J.S., Zurawski, S.M. and Lee, F.D. (1986) Science 232, 772–775.
[2] Howells, R.D., Kilpatrick, D.L., Bailey, L.C., Noe, M. and Udenfriend, S. (1986) Proc. Natl. Acad. Sci. USA 83, 1960–1963.
[3] Kilpatrick, D.L., Howells, R.D., Noe, M., Bailey, L.C. and Udenfriend, S. (1985) Proc. Natl. Acad. Sci. USA 82, 7467–7469.
[4] Kilpatrick, D.L. and Millette, C.F. (1986) Proc. Natl. Acad. Sci. USA 83, 5015–5018.
[5] Yoshikawa, K. and Aizawa, T. (1988) Biochem. Biophys. Res. Commun. 151, 664–671.
[6] Yoshikawa, K. and Aizawa, T. (1988) FEBS Lett. 237, 183–186.
[7] Chirgwin, J.M., Przbyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[8] Yoshikawa, K., Williams, C. and Sabol, S.L. (1984) J. Biol. Chem. 259, 14301–14308.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Yoshikawa, K. and Aizawa, T. (1988) Mol. Brain Res. 4, 87–96.
[11] Yoshikawa, K. and Sabol, S.L. (1986) Mol. Brain Res. 1, 75–83.
[12] Rosen, H., Douglass, J. and Herbert, E. (1984) J. Biol. Chem. 259, 14309–14313.
[13] Howells, R.D., Kilpatrick, D.L., Bhatt, R., Monahan, J.J., Poonian, M. and Udenfriend, S. (1984) Proc. Natl. Acad. Sci. USA 81, 7651–7655.
[14] Vilijn, M.-H., Vaysse, P.J.J., Zukin, R.S. and Kessler, J.A. (1988) Proc. Natl. Acad. Sci. USA 85, 6551–6555.
[15] Golub, E.S. (1982) Nature 299, 483.