Isolation and characterization of the mouse corticotropin–β-lipotropin precursor gene and a related pseudogene

Mitsue Notake, Takamasa Tobimatsu, Yumiko Watanabe, Hideo Takahashi, Masayoshi Mishina and Shosaku Numa*

Department of Medical Chemistry, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606, Japan

Received 11 April 1983

Two mouse genomic DNA sequences homologous with human corticotropin–β-lipotropin precursor gene sequences have been cloned. One of them represents the functional corticotropin–β-lipotropin precursor gene, which exhibits a structural organization similar to those of its bovine and human counterparts. The other represents a pseudogene that corresponds to the functional mouse gene sequence encoding the carboxy-terminal 143 amino acid residues (including corticotropin and β-lipotropin) and the 3′-untranslated region.

Mouse corticotropin–β-lipotropin precursor gene *Pseudogene* *Recombinant DNA* *DNA sequence* *Exon/intron* *Direct repeat*

1. INTRODUCTION

The primary structure of the common precursor of corticotropin (ACTH) and β-lipotropin (β-LPH) has been deduced from the nucleotide sequence of cloned DNA complementary to its mRNA [1] which was purified from the bovine pituitary intermediate lobe [2]. With the use of cDNA probes, the whole bovine and human ACTH-β-LPH precursor genes have been isolated, and their structural organizations have been elucidated [3–6]. Both genes consist of 3 mRNA-coding segments (designated as exon 1, exon 2 and exon 3 in the 5′ to 3′ direction on the message strand) divided by two large intervening sequences (designated as intron A and intron B in the same direction). Intron A is inserted within the segment corresponding to the 5′-untranslated region of the mRNA, and intron B within the protein-coding sequence near the signal peptide region. This investigation deals with the cloning and sequence analysis of the mouse ACTH–β-LPH precursor gene and a mouse pseudogene related to it.

2. MATERIALS AND METHODS

The gene library used, which was a kind gift from Dr T. Honjo and Dr T. Kataoka, was a collection of recombinant phage that carried mouse MPC11 myeloma DNA fragments generated by partial digestion with *Eco*RI and joined to the λ Charon 4A arms [7]. Screening of phage was performed as in [8]; the hybridization probes were labelled with [α-^{32}P]dCTP by nick-translation [9]. Cellular DNA was prepared by the procedure in [10] and total RNA by the procedure in [11]. Restriction endonuclease mapping and blot hybridization analysis were conducted as in [3] and S_1 nuclease mapping as in [12]. DNA sequencing was carried out by the procedure in [13]. Reagents were obtained as in [3].

* To whom correspondence should be addressed

Abbreviations: ACTH, corticotropin; β-LPH, β-lipotropin

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00

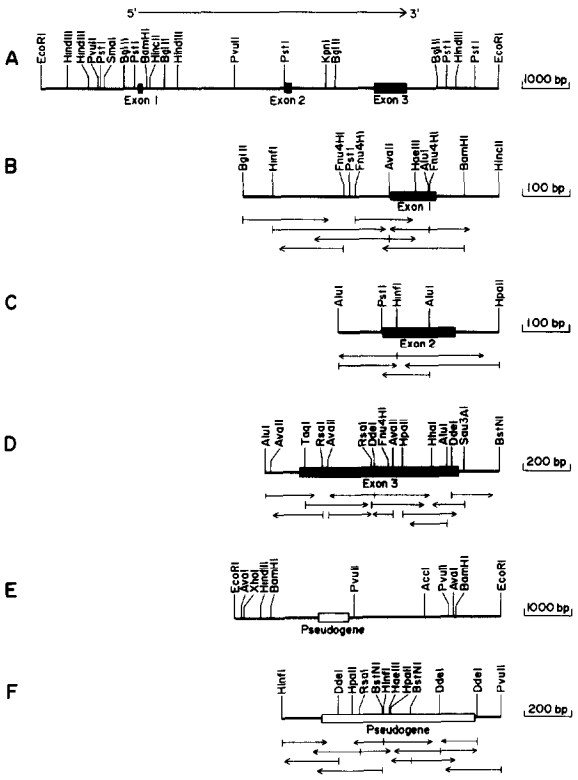
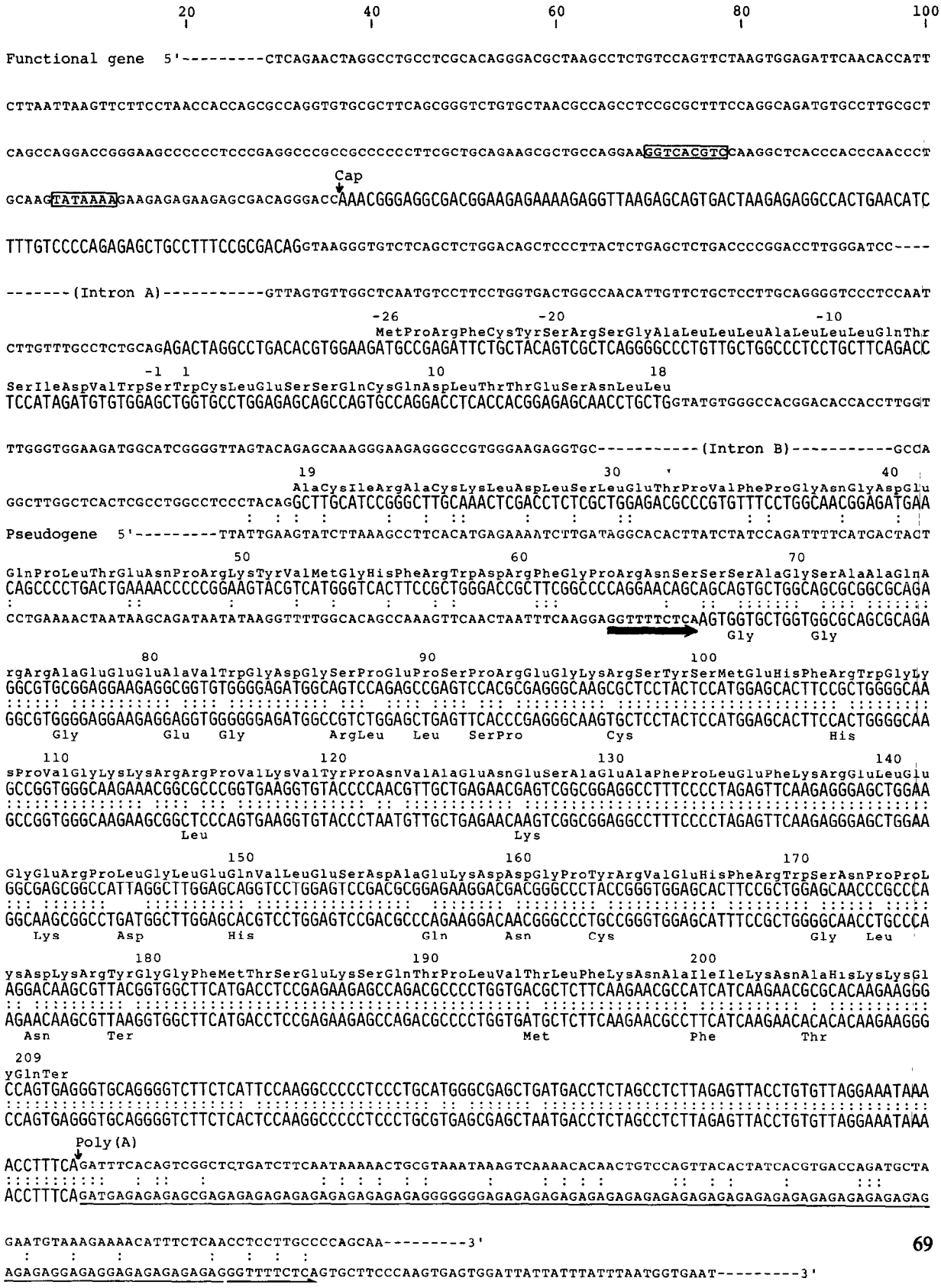


Fig.1. Restriction endonuclease mapping of cloned mouse genomic DNA segments containing the ACTH–β-LPH precursor gene and the pseudogene and sequencing strategy: (A) 9700 basepair *Eco*RI fragment containing the ACTH–β-LPH precursor gene; (B–D) portions of the 9700 basepair *Eco*RI fragment containing exon 1, exon 2 and exon 3, respectively; (E) 5600 basepair *Eco*RI fragment containing the pseudogene; (F) a portion of the 5600 basepair *Eco*RI fragment containing the sequence homologous with the functional gene sequence. The direction of transcription is from left to right. Scales are given in basepairs (bp) on the right side of each restriction map. For reference, the locations of the exons and the pseudogene are shown by closed and open boxes, respectively. Only the relevant restriction sites are displayed for: *Alu*I, *Fnu*4HI and *Hae*III (B); *Alu*I, *Bst*NI, *Fnu*4HI, *Hha*I and *Hpa*II (D); *Dde*I, *Hae*III and *Hpa*II (F). The horizontal arrows beneath the restriction maps in B–D and F indicate the direction and extent of sequence determinations.

3. RESULTS AND DISCUSSION

A mouse genomic DNA library [7] was screened for phage carrying ACTH–β-LPH precursor gene sequences by hybridization in situ at 60°C with ^{32}P-labelled human genomic DNA probes. Using the 657 basepair *Sma*I–*Sal*I fragment derived from exon 3 of the human ACTH–β-LPH precursor gene [4] as a probe, we isolated 41 hybridization-positive clones from ~5 × 10^6 plaques. These clones were rescreened by hybridization with the 172 basepair *Ava*I fragment containing most of exon 1 and its preceding region [4] or with the ~790 basepair *Bam*HI–*Hin*fI fragment containing exon 2 and its flanking regions [4]. Ten (first type) of the 41 clones hybridized with both the exon 1 and exon 2 probes, while the remaining 31 clones (second type) did not hybridize with either probe. Blot hybridization analysis of DNA prepared from 2 clones of the first type and 3 clones of the second type showed that the former clones carried a common *Eco*RI fragment of ~9700 basepairs, while

Fig.2. Nucleotide sequence of the exons and their surrounding regions of the mouse ACTH–β-LPH precursor gene aligned with that of the pseudogene. The nucleotide sequence of the message strand of the functional gene, together with the deduced amino acid sequence, is shown; the sequences of the exons and their surrounding regions are represented by large and small capital letters, respectively; the amino acid residues are numbered beginning with the tryptophan residue specified by the 27th codon, and the preceding residues constituting the signal peptide [14] are indicated by negative numbers. The exon/intron junctions are positioned according to the GT/AG rule [15]. The putative sites of capping and poly(A) addition are indicated. The TATA box and the CAAT box are enclosed with boxes. The nucleotide sequence of the pseudogene, aligned with that of the functional gene, is shown beneath the latter; the homologous pseudogene sequence and its flanking sequences are represented by large and small capital letters, respectively. The colons indicate identical nucleotides in the two sequences. The amino acid differences anticipated from the pseudogene sequence are shown for the homologous region; Ter denotes a translational termination codon, and the absence of an amino acid indicates that an identical residue is specified by the functional gene and the pseudogene. The pseudogene segment composed mainly of repeated GA dinucleotides is underlined. The 10 basepair direct repeats are indicated by thick arrows beneath the sequences.


Functional gene 5'---------CTCAGAACTAGGCCTGCCTCGCACAGGGACGCTAAGCCTCTGTCCAGTTCTAAGTGGAGATTCAACACCATT
CTTAATTAAGTTCTTCCTAACCACCAGCGCCAGGTGTGCGCTTCAGCGGGTCTGTGCTAACGCCAGCCTCCGCGCTTTCCAGGCAGATGTGCCTTGCGCT
CAGCCAGGACCGGGAAGCCCCCCTCCCGAGGCCCGCCGCCCCCCTTCGCTGCAGAAGCGCTGCCAGGAAGGTCACGTCCAAGGCTCACCCACCCAACCCT
Cap
GCAAGTATAAAAGAAGAGAGAAGAGCGACAGGGACCAAACGGGAGGCGACGGAAGAGAAAAGAGGTTAAGAGCAGTGACTAAGAGAGGCCACTGAACATC
TTTGTCCCCAGAGAGCTGCCTTTCCGCGACAGGTAAGGGTGTCTCAGCTCTGGACAGCTCCCTTACTCTGAGCTCTGACCCCGGACCTTGGGATCC----
-------(Intron A)----------GTTAGTGTTGGCTCAATGTCCTTCCTGGTGACTGGCCAACATTGTTCTGCTCCTTGCAGGGGTCCCTCCAAT
-26 -20 -10
MetProArgPheCysTyrSerArgSerGlyAlaLeuLeuLeuAlaLeuLeuLeuGlnThr
CTTGTTTGCCTCTGCAGAGACTAGGCCTGACACGTGGAAGATGCCGAGATTCTGCTACAGTCGCTCAGGGGCCCTGTTGCTGGCCCTCCTGCTTCAGACC
-1 1 10 18
SerIleAspValTrpSerTrpCysLeuGluSerSerGlnCysGlnAspLeuThrThrGluSerAsnLeuLeu
TCCATAGATGTGTGGAGCTGGTGCCTGGAGAGCAGCCAGTGCCAGGACCTCACCACGGAGAGCAACCTGCTGGTATGTGGGCCACGGACACCACCTTGGT
TTGGGTGGAAGATGGCATCGGGGTTAGTACAGAGCAAAGGGAAGAGGGCCGTGGGAAGAGGTGC----------(Intron B)----------GCCA
19 30 40
AlaCysIleArgAlaCysLysLeuAspLeuSerLeuGluThrProValPheProGlyAsnGlyAspGlu
GGCTTGGCTCACTCGCCTGGCCTCCCTACAGGCTTGCATCCGGGCTTGCAAACTCGACCTCTCGCTGGAGACGCCCGTGTTTCCTGGCAACGGAGATGAA
Pseudogene 5'---------TTATTGAAGTATCTTAAAGCCTTCACATGAGAAAATCTTGATAGGCACACTTATCTATCCAGATTTTCATGACTACT
50 60 70
GlnProLeuThrGluAsnProArgLysTyrValMetGlyHisPheArgTrpAspArgPheGlyProArgAsnSerSerSerAlaGlySerAlaAlaGlnA
CAGCCCCTGACTGAAAACCCCCGGAAGTACGTCATGGGTCACTTCCGCTGGGACCGCTTCGGCCCCAGGAACAGCAGCAGTGCTGGCAGCGCGGCGCAGA
CCTGAAAACTAATAAGCAGATAATATAAGGTTTTGGCACAGCCAAAGTTCAACTAATTTCAAGGAGGTTTTCTCAAGTGGTGCTGGTGGCGCAGCGCAGA
Gly Gly
80 90 100
rgArgAlaGluGluGluAlaValTrpGlyAspGlySerProGluProSerProArgGluGlyLysArgSerTyrSerMetGluHisPheArgTrpGlyLy
GGCGTGCGGAGGAAGAGGCGGTGTGGGGAGATGGCAGTCCAGAGCCGAGTCCACGCGAGGGCAAGCGCTCCTACTCCATGGAGCACTTCCGCTGGGGCAA
GGCGTGGGGAGGAAGAGGAGGTGGGGGGAGATGGCCGTCTGGAGCTGAGTTCACCCGAGGGCAAGTGCTCCTACTCCATGGAGCACTTCCACTGGGGCAA
Gly Glu Gly ArgLeu Leu SerPro Cys His
110 120 130 140
sProValGlyLysLysArgArgProValLysValTyrProAsnValAlaGluAsnGluSerAlaGluAlaPheProLeuGluPheLysArgGluLeuGlu
GCCGGTGGGCAAGAAACGGCGCCCGGTGAAGGTGTACCCCAACGTTGCTGAGAACGAGTCGGCGGAGGCCTTTCCCCTAGAGTTCAAGAGGGAGCTGGAA
GCCGGTGGGCAAGAAGCGGCTCCCAGTGAAGGTGTACCCTAATGTTGCTGAGAACAAGTCGGCGGAGGCCTTTCCCCTAGAGTTCAAGAGGGAGCTGGAA
Leu Lys
150 160 170
GlyGluArgProLeuGlyLeuGluGlnValLeuGluSerAspAlaGluLysAspAspGlyProTyrArgValGluHisPheArgTrpSerAsnProProL
GGCGAGCGGCCATTAGGCTTGGAGCAGGTCCTGGAGTCCGACGCGGAGAAGGACGACGGGCCCTACCGGGTGGAGCACTTCCGCTGGAGCAACCCGCCCA
GGCAAGCGGCCTGATGGCTTGGAGCACGTCCTGGAGTCCGACGCCCAGAAGGACAACGGGCCCTGCCGGGTGGAGCATTTCCGCTGGGGCAACCTGCCCA
Lys Asp His Gln Asn Cys Gly Leu
180 190 200
ysAspLysArgTyrGlyGlyPheMetThrSerGluLysSerGlnThrProLeuValThrLeuPheLysAsnAlaIleIleLysAsnAlaHisLysLysGl
AGGACAAGCGTTACGGTGGCTTCATGACCTCCGAGAAGAGCCAGACGCCCCTGGTGACGCTCTTCAAGAACGCCATCATCAAGAACGCGCACAAGAAGGG
AGAACAAGCGTTAAGGTGGCTTCATGACCTCCGAGAAGAGCCAGACGCCCCTGGTGATGCTCTTCAAGAACGCCTTCATCAAGAACACACACAAGAAGGG
Asn Ter Met Phe Thr
209
yGlnTer
CCAGTGAGGGTGCAGGGGTCTTCTCATTCCAAGGCCCCCTCCCTGCATGGGCGAGCTGATGACCTCTAGCCTCTTAGAGTTACCTGTGTTAGGAAATAAA
CCAGTGAGGGTGCAGGGGTCTTCTCACTCCAAGGCCCCCTCCCTGCGTGAGCGAGCTAATGACCTCTAGCCTCTTAGAGTTACCTGTGTTAGGAAATAAA
Poly(A)
ACCTTTCAGATTTCACAGTCGGCTCTGATCTTCAATAAAAACTGCGTAAATAAAGTCAAAACACAACTGTCCAGTTACACTATCACGTGACCAGATGCTA
ACCTTTCAGATGAGAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GAATGTAAAGAAAACATTTCTCAACCTCCTTGCCCCAGCAA---------3'
AGAGAGGAGAGGAGAGAGAGAGAGGGTTTTCTCAGTGCTTCCCAAGTGAGTGGATTATTATTTATTTAATGGTGAAT---------3'

the latter clones harboured a common *Eco*RI fragment of ~5600 basepairs. The sizes of these *Eco*RI fragments agree with those of the two hybridization-positive *Eco*RI fragments detected by blot hybridization analysis of cellular DNA (isolated from BALB/c mouse livers and cultures of the ACTH-producing mouse pituitary tumor cell line AtT-20/D-16v). The two *Eco*RI fragments from the recombinant phage were subcloned in plasmid pBR322 for further analysis.

Fig.1A and E show restriction maps of the cloned 9700 basepair and 5600 basepair *Eco*RI fragments, respectively. The hybridization-positive DNA segments and their surrounding regions were subjected to nucleotide sequence analysis according to the strategy illustrated in fig.1B–D and F. Comparison of the determined sequences with the bovine and human ACTH–β-LPH precursor gene sequences [3,4] clearly indicated that the homologous mouse sequence contained in the 9700 basepair *Eco*RI fragment represents the functional ACTH–β-LPH precursor gene, while that in the 5600 basepair *Eco*RI fragment represents a pseudogene. The two mouse sequences are aligned in fig.2.

The functional mouse ACTH–β-LPH precursor gene contains two introns as identified by comparison with its bovine and human counterparts [3,4] and with the cDNA for the mouse ACTH–β-LPH precursor reported in [16]. Intron A, ~3000 basepairs long, interrupts the segment encoding the 5′-untranslated region between 23 and 24 basepairs upstream of the translational initiation site. Intron B, ~1700 basepairs long, is inserted in the protein-coding region, separating the segment encoding the amino-terminal 44 amino acids including the signal peptide from that encoding the remaining 191 amino acids and the 3′-untranslated region. Thus, the mouse ACTH–β-LPH precursor gene, consisting of 3 exons, exhibits a structural organization similar to those of its bovine and human counterparts. The capping site and the poly(A) addition site of the mouse gene have tentatively been assigned by comparison with the bovine and human genes [3,4] and with the mouse cDNA [16], respectively. The assignment of the capping site is supported by S_1 nuclease mapping using total RNA from cultured AtT-20/D-16v cells and the 162 basepair *Pst*I–*Alu*I fragment (5′-end-labelled at the *Alu*I site; see fig.1B). A putative TATA box [17] and a putative CAAT box [18] are found 25 basepairs and 59 basepairs upstream of the capping site assigned, respectively. Comparison of the nucleotide sequence of the mouse ACTH–β-LPH precursor gene with that of the mouse cDNA [16], cloned with the use of the mRNA from the cell line AtT-20/D-16v, shows some nucleotide differences, two of which result in amino acid replacements (Ser instead of Arg at residue 19; His instead of Gln at residue 150). The observed nucleotide differences may be due to possible heterogeneity in the mouse ACTH–β-LPH precursor gene or to possible errors introduced during DNA cloning.

The pseudogene exhibits a nucleotide sequence highly homologous with the 3′-terminal 533 basepairs of the functional gene, which constitute a large portion of exon 3 including the coding regions for ACTH (residues 98–136) and β-LPH (residues 139–209). The degree of homology is 92%, and neither insertions nor deletions are found. Many of the nucleotide substitutions, however, would result in amino acid replacements, for example, the replacement of Arg (residue 97) by Cys, which would abolish the Lys–Arg pair serving as a signal for proteolytic processing [19] to produce ACTH, and the introduction of a translational termination codon at position 179 (instead of Tyr), which would lead to premature termination at the amino terminus of β-endorphin (residues 179–209). Thus, the pseudogene cannot encode a functional precursor protein like the ACTH–β-LPH precursor.

The sequence homology between the functional gene and the pseudogene ends, on the 5′-side, 94 basepairs upstream of the 5′-end of the ACTH-coding region. No evidence for the presence of homologous sequences in the further upstream region was provided by blot hybridization analysis of the phage DNA extending up to ~7000 basepairs upstream of the 5′-end of the homologous pseudogene sequence; the 466 basepair *Bgl*II–*Bam*HI fragment and the ~1100 basepair *Pst*I–*Bgl*II fragment, containing exon 1 and exon 2 of the functional mouse gene, respectively (see fig.1A), were used as hybridization probes. On the 3′-side, the sequence homology ends at the putative poly(A) addition site. In the pseudogene this site is followed by a 116 basepair segment composed mainly of repeated GA

dinucleotides. The 649 basepair pseudogene sequence including the GA repeats, is flanked on both sides by direct repeats of the 10 basepair sequence GGTTTTCTCA. This suggests the possibility that the pseudogene may have been generated by the formation of an aberrant transcript of the functional gene followed by the insertion of its DNA copy into the mouse genome, although a typical poly(A) tail is not found. This view is supported by our in vivo expression studies suggesting the aberrant initiation of transcription of the human ACTH–β-LPH precursor gene around the site corresponding to the 5′-end of the homologous pseudogene sequence ([20]; unpublished). An RNA intermediate may be involved in the generation of some pseudogenes [21–26]. In [27] the presence of a mouse pseudogene related to the ACTH–β-LPH precursor gene was mentioned.

ACKNOWLEDGEMENTS

We thank Dr T. Honjo and Dr T. Kataoka for the mouse genomic DNA library. 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] Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S. (1979) Nature 278, 423–427.
[2] Kita, T., Inoue, A., Nakanishi, S. and Numa, S. (1979) Eur. J. Biochem. 93, 213–220.
[3] Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429–438.
[4] Takahashi, H., Teranishi, Y., Nakanishi, S. and Numa, S. (1981) FEBS Lett. 135, 97–102.
[5] Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) Nature 297, 335–339.
[6] Whitfeld, P.L., Seeburg, P.H. and Shine, J. (1982) DNA 1, 133–143.
[7] Kataoka, T., Miyata, T. and Honjo, T. (1981) Cell 23, 357–368.
[8] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[9] Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel, R. (1978) Proc. Natl. Acad. Sci. USA 75, 1299–1303.
[10] Polsky, F., Edgell, M.H., Seidman, J.G. and Leder, P. (1978) Anal. Biochem. 87, 397–410.
[11] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[12] Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721–732.
[13] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[14] Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67, 852–862.
[15] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[16] Uhler, M. and Herbert, E. (1983) J. Biol. Chem. 258, 257–261.
[17] Goldberg, M. (1979) PhD thesis, Stanford University.
[18] Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127–142.
[19] Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980) Ann. NY Acad. Sci. 343, 1–16.
[20] Mishina, M., Kurosaki, T., Yamamoto, T., Notake, M., Masu, M. and Numa, S. (1982) EMBO J. 1, 1533–1538.
[21] Nishioka, Y., Leder, A. and Leder, P. (1980) Proc. Natl. Acad. Sci. USA 77, 2806–2809.
[22] Vanin, E.F., Goldberg, G.I., Tucker, P.W. and Smithies, O. (1980) Nature 286, 222–226.
[23] Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Nature 296, 321–325.
[24] Wilde, C.D., Crowther, C.E. and Cowan, N.J. (1982) Science 217, 549–552.
[25] Karin, M. and Richards, R.I. (1982) Nature 299, 797–803.
[26] Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330–335.
[27] Herbert, E. and Uhler, M. (1982) Cell 30, 1–2.