9. Morris, V. M., Medeiros, E., Ringold, G. M., Bishop, J. M. & Varmus, H. E. *J. molec. Biol.* **114**, 73–92 (1977).
10. Michelides, R., Van Deemter, L., Nusse, R. & Van Nie, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2368–2372 (1978).
11. Michalides, R. & Schlom, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4635–4639 (1975).
12. Parks, W. P. & Scolnick, E. M. *Virology* **55**, 163–173 (1973).
13. Varmus, H. E., Stavnezer, J., Medeiros, E. & Bishop, J. M. in *Comp. Leuk. Res.* (eds Ito, Y. & Dutcher, R. M.) 451–461 (University of Tokyo Press, Basel, 1975).
14. Botchan, M., Topp, W. & Sambrook, J. *Cell* **9**, 269–287 (1976).
15. Britten, R. J., Cetta, A. & Davidson, E. H. *Cell* **15**, 1175–1186 (1978).
16. Cohen, J. C., Shank, P. R., Morris, V. L., Cardiff, R. & Varmus, H. E. *Cell* (in the press).
17. Southern, E. M. *J. molec. Biol.* **98**, 502–517 (1975).
18. Shank, P. R., Cohen, J. C., Varmus, H. E., Yamamoto, K. R. & Ringold, G. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2112–2116 (1978).
19. Ringold, G. M., Shank, P. R., Varmus, H. E., Ring, J. & Yamamoto, K. R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
20. Rice, N. R. & Strauss, N. A. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3546–3550 (1973).
21. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (eds) *DNA Insertion Elements, Plasmids, and Episomes* (Cold Spring Harbor Laboratory, New York, 1977).
22. Staats, J. *Cancer Res.* **36**, 4333–4777 (1976).
23. Smith, G. H. *J. natn. Cancer Inst.* **36**, 685–694 (1966).
24. Morris V. L. *et al. Virology* **92**, 46–55 (1979).
25. Benveniste, R. E. & Todaro, G. J. *Nature* **252**, 456–459 (1974).
26. Benveniste, R. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4090–4094 (1975).
27. Hughes, S., Payvar, F., Spector, D., Schimke, R. T., Robinson, H. L., Bishop, J. M. & Varmus, H. E. (in preparation).
28. Astrin, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5941–5945 (1978).
29. Steffen, D. & Weinberg, R. A. *Cell* **15**, 1003–1010 (1978).
30. Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. & Rowe, W. P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 167–171 (1974).
31. Benveniste, R. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4513–4518 (1974).
32. Benveniste, R. E. & Todaro, G. J. *Nature* **261**, 101–108 (1976).
33. Jaenisch, R. *Cell* **12**, 691–696 (1977).
34. Van Nie, R. & De Moes, J. *Int. J. Cancer* **20**, 588–594 (1977).
35. Michalides, R., Vlahakis, G. & Schlom, J. *Int. J. Cancer* **18**, 105–115 (1976).
36. Cooper, G. M. & Temin, H. M. *J. Virol.* **14**, 1132–1141 (1974).
37. Shank, P. R. *et al. Cell* **15**, 1383–1395 (1978).
38. Hughes, S. H. *et al. Cell* **15**, 1397–1410 (1978).
39. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. *J. Virol.* **28**, 810–818 (1978).
40. Sabran, J. L. *et al. J. Virol.* **29**, 170–178 (1979).
41. Greene, P. J. *et al. Nucleic Acids Res.* **5**, 2373–2380 (1978).
42. Denhardt, D. *Biochem. biophys. Res. Commun.* **23**, 641–646 (1966).
43. Swanstrom, R. & Shank, P. R. *Analyt. Biochem.* **86**, 184–192 (1978).
44. Gardner, M. B. *Curr. Topics Microb. Immun.* **79**, 215–259 (1978).

Nucleotide sequence of cloned cDNA for bovine corticotropin–β-lipotropin precursor

Shigetada Nakanishi*, Akira Inoue*, Toru Kita*, Masahiro Nakamura*, Annie C. Y. Chang†, Stanley N. Cohen† & Shosaku Numa*

* Department of Medical Chemistry, Kyoto University Faculty of Medicine, Kyoto 606, Japan
and
† Departments of Genetics and Medicine, Stanford University School of Medicine, Stanford, California 94305

The nucleotide sequence of a 1,091-base pair cloned cDNA insert encoding bovine corticotropin–β-lipotropin precursor mRNA is reported. The corresponding amino acid sequence indicates that the precursor protein consists of repetitive units and includes a third melanotropin sequence in its cryptic portion. Pairs of lysine and arginine residues separate the component peptides of the precursor.

THE pituitary hormones corticotropin (ACTH) and β-lipotropin (β-LPH) and their gene provide a useful system for studying the structure and regulation of hormonally regulated genes. These two hormones, which are formed from a large common precursor protein[1–8], contain small component peptides with biological activity; α-melanotropin (α-MSH) and corticotropin-like intermediate-lobe peptide (CLIP) are derived from ACTH[9], whereas γ-lipotropin (γ-LPH), β-melanotropin (β-MSH), endorphins and methionine-enkephalin are elaborated from β-LPH[10,11]. The intracellular level of the mRNA coding for the common precursor protein is depressed by glucocorticoids, which probably act at the transcriptional level by means of the glucocorticoid receptor[12,13]. As ACTH and β-LPH together account for only one-third to one-half of the molecular weight of the precursor protein, there has been considerable interest in and speculation about the primary structure and possible biological functions of the remaining portion.

Recently developed techniques of DNA cloning[14] and nucleotide sequence analysis[15,16] together provide a useful approach to an understanding of the primary structure of eukaryotic genes and the regulation of their expression. Previously[17], we described the construction of bacterial plasmids that contain the nucleotide sequence for the bovine ACTH–β-LPH precursor and presented the DNA sequence coding for ACTH and a portion of β-LPH. Here, we report the complete nucleotide sequence of a 1,091-base pair cloned cDNA insert that contains a duplex copy of the ACTH–β-LPH precursor mRNA. Our results define the precise locations of ACTH and β-LPH in the precursor protein and predict the amino acid sequence of its remaining portion. They further indicate that the precursor protein consists of several repetitive units separated by pairs of basic amino acid residues, lysine and arginine, and that a third melanotropin sequence is located within the cryptic portion of the molecule.

Sequencing strategy

The bacterial plasmid pSNAC20 was constructed by inserting *in vitro*-synthesised double-stranded cDNA for the bovine ACTH–β-LPH precursor into the *Pst*I endonuclease cleavage site of the plasmid pBR322 with the use of poly(dG)·poly(dC) homopolymeric extensions[17]. Nucleotide sequence analysis of the inserted cDNA was carried out by the procedure of Maxam and Gilbert[15]; the sequence map is shown in Fig. 1. The restriction fragments used were obtained either from intact pSNAC20 or from the cDNA insert (including poly(dG)·poly(dC) tails), which was isolated by cleavage of the plasmid at the regenerated *Pst*I sites[17].

For accurate prediction of the amino acid sequence of the cryptic portion of the ACTH–β-LPH precursor protein, the nucleotide sequence was determined on both strands of the cDNA for all but 33 residues corresponding to the 5′ end and 47 residues corresponding to the 3′ end of the mRNA; for these regions this was technically difficult, but the sequence data were reliable. In addition, care was taken to overlap the sequence data derived from different endonuclease-generated fragments. These procedures provided a useful verification of the sequence. Furthermore, some of the endonuclease-generated fragments were analysed for nucleotide sequence in duplicate or in triplicate. Thus, at least three sets of sequence data were obtained for all segments of the cloned cDNA, thus eliminating any ambiguity resulting from closely lying bands of chemically degraded DNA, which were occasionally observed on electrophoretic analysis.

Another potential problem in the use of cloned cDNA for sequence analysis of mRNA is a possible error caused by deletion, addition or rearrangement of nucleotides during *in vitro* cDNA synthesis or DNA cloning itself[18–20]. Thus, an independent method, such as analysis of protein sequence or mRNA sequence, should preferably be used to verify that the

© Macmillan Journals Ltd 1979

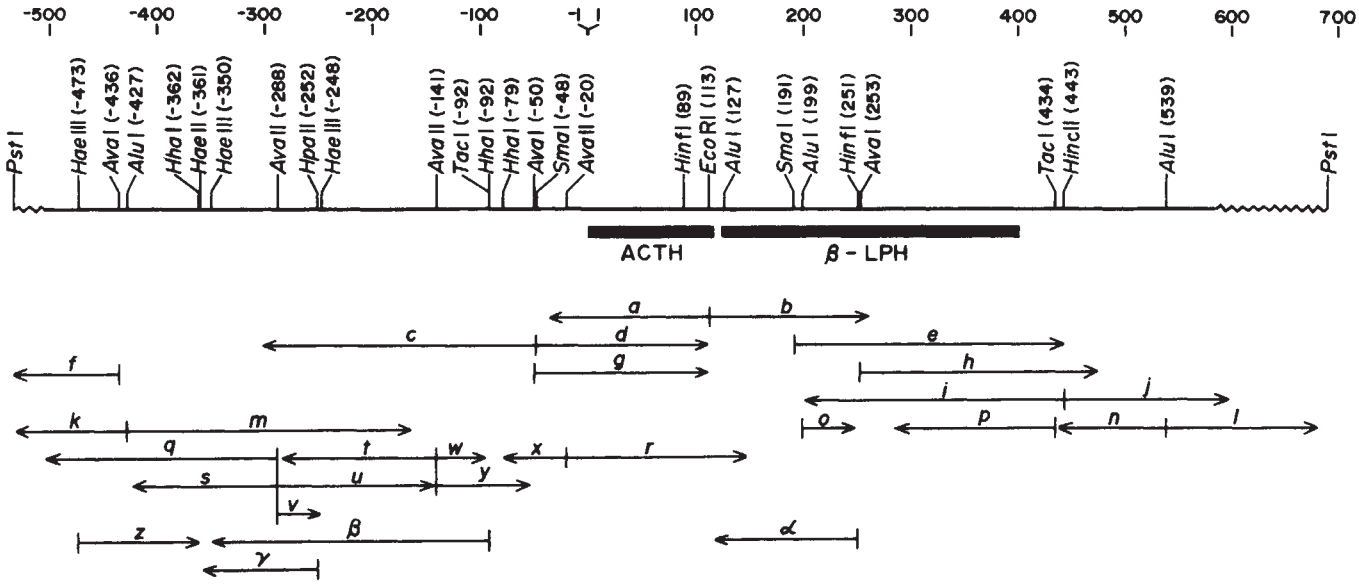

Fig. 1 Sequence map for the cDNA insert in the pSNAC20 plasmid. Nucleotide residues are numbered in the direction from 5′ to 3′ in the message strand, beginning with the first residue in the coding region for ACTH, and the nucleotides on the 5′ side of residue 1 are indicated by negative numbers. The wavy lines represent the poly(dG)·poly(dC) tails of the cDNA insert (27 and ~100 base pairs long on the left and right sides, respectively). The coding regions for ACTH and β-LPH are indicated by closed boxes. Only the restriction sites used for sequencing are shown and are identified by numbers indicating the 5′-terminal nucleotide generated by cleavage. DNA sequence analysis was carried out by the procedure of Maxam and Gilbert[15]. The direction and extent of the sequence determinations are indicated by horizontal arrows *a–z* and *α–γ*. The 5′-end-labelled restriction fragments used for sequencing, which were derived either directly from intact pSNAC20 or from the cDNA insert obtained by cleavage of the plasmid with *Pst*I followed by electrophoretic isolation, were prepared as follows (the letters in parentheses refer to the horizontal arrows; the labelled site is designated by an asterisk, and the number indicates the fragment chain length in base pairs[52]; the chain lengths of fragments including the poly(dG):poly(dC) tail on the right side are approximate): *a, b*, Cleavage of pSNAC20 with *Eco*RI, labelling, cleavage with *Pst*I, isolation of **Eco*RI-*Pst*I-645 (*a*) and **Eco*RI-*Pst*I-570 (*b*). *c–e*, Cleavage of pSNAC20 with *Sma*I, labelling, cleavage with *Eco*RI, isolation of **Sma*I-*Eco*RI-1, 241 (*c*), **Sma*I-*Eco*RI-160 (*d*) and **Sma*I-*Eco*RI-4, 100 (*e*). *f–h*, Cleavage of pSNAC20 with *Ava*I, labelling, cleavage with *Pvu*I, *Eco*RI or *Pst*I, isolation of **Ava*I-*Pvu*I-223 (*f*), **Ava*I-*Eco*RI-162 (*g*) and **Ava*I-*Pst*I-430 (*h*). *i, j*, Cleavage of pSNAC20 with *Hinc*II, labelling, cleavage with *Pst*I, isolation of **Hinc*II-*Pst*I-975 (*i*) and **Hinc*II-*Pst*I-240 (*j*). *k–p*, Cleavage of the cDNA insert with *Alu*I, labelling, isolation of **Alu*I-*Pst*I-106 (*k*), **Alu*I-*Pst*I-150 (*l*), **Alu*I-**Alu*I-553 and **Alu*I-**Alu*I-340 (the *Pst*I site was not significantly labelled), cleavage of **Alu*I-**Alu*I-553 with *Sma*I, isolation of **Alu*I-*Sma*I-379 (*m*), cleavage of **Alu*I-**Alu*I-340 with *Tac*I, isolation of **Alu*I-*Tac*I-105 (*n*) and **Alu*I-*Tac*I-235, labelling of **Alu*I-*Tac*I-235, cleavage with *Hinf*I, isolation of **Alu*I-*Hinf*I-52 (*o*) and **Tac*I-*Hinf*I-183 (*p*). *q–y*, Cleavage of the cDNA insert with *Ava*II, labelling, isolation of **Ava*II-*Pst*I-245 (*q*), **Ava*II-*Pst*I-710 (*r*), **Ava*II-**Ava*II-147 and **Ava*II-**Ava*II-121 (in one experiment, **Ava*II-*Pst*I-245 was cleaved with *Alu*I, and **Ava*II-*Alu*I-139 (*s*) was isolated), alkali treatment[53] of **Ava*II-**Ava*II-147, isolation of single-stranded **Ava*II-*Ava*II-147 (*t, u*) (in one experiment, **Ava*II-**Ava*II-147 was cleaved with *Hae*III, and **Ava*II-*Hae*III-40 (*v*) was isolated), cleavage of **Ava*II-**Ava*II-121 either with *Hha*I or with *Sma*I, isolation of **Ava*II-*Hha*I-49 (*w*), **Ava*II-*Hha*I-59 (*x*) and **Ava*II-*Sma*I-93 (*y*). *z*, Cleavage of the cDNA insert with *Ava*II, isolation followed by cleavage with *Hae*III of *Ava*II-*Pst*I-245, isolation followed by labelling of *Hae*III-*Hae*III-123, cleavage with *Hae*II, isolation of **Hae*III-*Hae*II-112. *α*, Cleavage of the cDNA insert with *Ava*II, isolation followed by cleavage with *Hinf*I of *Ava*II-*Pst*I-710, isolation followed by labelling of *Hinf*I-*Hinf*I-162, cleavage with *Eco*RI, isolation of **Hinf*I-*Eco*RI-138. *β*, Cleavage of the cDNA insert with *Eco*RI, isolation followed by cleavage with *Tac*I of *Eco*RI-*Pst*I-645, isolation followed by labelling of *Tac*I-*Pst*I-441, cleavage with *Hha*I, isolation of **Tac*I-*Hha*I-270. *γ*, Cleavage of the cDNA insert with *Tac*I, isolation followed by cleavage with *Hpa*II of *Tac*I-*Pst*I-441, labelling, cleavage with *Hha*I, isolation of **Hpa*II-*Hha*I-110. Cleavage with restriction endonucleases was carried out in the conditions described by the vendors with optimal amounts of the enzymes. Separation of the restriction fragments was carried out by electrophoresis on either 5% or 8.3% polyacrylamide slab gel in 50 mM Tris-borate containing 1 mM EDTA (*p*H 8.3), except that alkali-treated **Ava*II-**Ava*II-147 was electrophoresed on 15% gel for strand separation. Fractionated DNA was extracted from the gel by maceration. Treatment of DNA with 7 μg of bacterial alkaline phosphatase (Worthington) was carried out at 55 °C for 1 h in 20 mM Tris-HCl (*p*H 8.3). Labelling of DNA was carried out at 37 °C for 1 h with 200–250 μCi [γ-^{32}P]ATP (~7,000 Ci mmol^{-1}, New England Nuclear) and 0.5 μg of T4 polynucleotide kinase (Biogenics Research Corporation) in a reaction mixture containing 50 mM Tris-HCl (*p*H 9.5), 10 mM $MgCl_2$, 5 mM dithiothreitol and 0.75 mM spermidine.

cloned cDNA we have studied represents an exact copy of the mRNA. However, the fidelity of cDNA cloning in bacterial plasmids has been amply established for several mRNA species[21–26]. We have also obtained evidence indicating that the coding regions for the known segments of the precursor protein (that is, the segments containing ACTH and β-LPH) are faithfully represented in the cloned cDNA (see below). In any case, the conclusions reported here result from study of the cDNA copy of a single mRNA molecule and heterogeneity of the ACTH–β-LPH precursor mRNA cannot be excluded.

Sequence of ACTH–β-LPH precursor mRNA and assignment of protein sequence

Figure 2 gives the sequence of 1,091 nucleotides comprising the cDNA inserted in the bacterial plasmid pSNAC20 (excluding poly(dG)·poly(dC) tails). Our previous work has demonstrated that the sequence of nucleotide residues 1–207 corresponds precisely to the amino acid sequences determined for bovine ACTH and the first portion of bovine β-LPH (28 amino acid residues) plus the two amino acid residues Lys-Arg connecting the two peptides[17]. The nucleotide sequence we have determined for the remaining portion of β-LPH indicates the presence of two additional amino acid residues, Ala-Glu, between positions 35 and 36 of the generally accepted structure of bovine β-LPH[27]. Thus, our results are consistent with a structure of bovine β-LPH that contains 93 amino acid residues, as reported by Pankov[28]; however, our finding that the residue at position 16 is Gln agrees with the assignment at this position made by Li *et al.*[27].

The nucleotide residues encoding the glutamine at the carboxyl terminus of β-LPH are followed by the translational termination codon UGA. This observation indicates that there is no peptide extension beyond the carboxyl terminus of β-LPH and that the 3′ noncoding region of the mRNA is composed of an untranslated region of 175 nucleotides plus a poly(A) terminal sequence. Another significant finding is that the amino terminus of ACTH, like the carboxyl terminus, is connected to the adjacent peptide by a pair of basic amino acid residues, Lys-Arg. Thus, the formation of ACTH and β-LPH from their precursor seems to involve proteolytic processing similar to that observed in the conversion of proinsulin to insulin[29] and proparathyroid hormone to parathyroid hormone[30,31].

In assigning the amino acid sequence of the cryptic portion of the ACTH–β-LPH precursor, the reading frame of amino acid codons was determined from the frame corresponding to the amino acid sequences of ACTH and β-LPH. However, the translational initiation site on this mRNA cannot be located definitively without knowing the amino acid sequence of the amino-terminal region of the precursor protein. A possible location for the translational initiation site is the methionine codon AUG at positions −393 to −391. On the basis of this assignment, the molecular weight of the ACTH–β-LPH precursor protein is calculated to be 29,259, a smaller value than the previously estimated molecular weight of the cell-free translation product of bovine ACTH–β-LPH precursor mRNA (35,000 (ref. 1) or 41,000 (ref. 2) depending on the electrophoresis system used). However, the relatively large number of basic amino acids (lysine and arginine) encoded by this mRNA (Fig. 2) or possible glycosylation of the cell-free trans-

©1979 Nature Publishing Group

lation product may result in overestimation of the molecular weight of the protein by electrophoretic analysis[32]. The assignment of the translational initiation site at the methionine codon mentioned above seems plausible because the first 20 amino acid residues starting with this putative initiative methionine (residue −131), which are followed by the characteristic sequence Ser-Met-Glu---- (see below), include a large number of hydrophobic amino acids (13 nonpolar residues including 7 leucine residues). A high content of hydrophobic amino acids is known to be present in the signal peptides of other proteins destined for secretion[33–36]. On the other hand, if the previous estimation of the molecular weight of the precursor protein is correct, the translational initiation site would be located 1,125–1,295 nucleotides (950–1,120 nucleotides for the coding region plus the 175 noncoding nucleotides we have identified at the 3′ end) away from the first residue of the poly(A) sequence of the mRNA and therefore would be absent from the cloned cDNA segment we have analysed. In Fig. 2, we have assigned amino acid residues up to the putative initiative methionine, although it remains possible that the region of the mRNA proximal to this site is translated.

Computer analysis of the amino acid sequence determined from the pSNAC20 clone indicates that the sequence of amino acid residues −55 to −44 is strikingly similar to the sequences of α-MSH and β-MSH, which are known to have common structural and biological features[37] (Fig. 3). This segment of the precursor protein shares with the known MSHs tyrosine and methionine residues as well as the characteristic tetrapeptide sequence His-Phe-Arg-Trp, which is required for melanotropic activity, at equivalent positions[38]. Moreover, the peptide fragment discovered by our DNA sequence analysis is flanked on both sides by a pair of basic amino acid residues. This suggests that the new peptide, like α-MSH or β-MSH, can be formed by proteolytic processing of the precursor; by analogy with α-MSH, it is possible that phenylalanine amide is formed at the carboxyl terminus of this peptide during processing. In view of the remarkable structural homology between the newly found peptide fragment and α-MSH or β-MSH, we propose to name the peptide γ-melanotropin (γ-MSH). Another peptide fragment within the precursor protein that exhibits some structural similarity to the MSHs is that containing amino acid residues −111 to −105 (Fig. 3). Note that this fragment contains

```
5'----AAGAGAACGAAGGGAAGAAAAGUGACCGAGAGGCCGCUGAACAUCCUCGCCCCGGCGCAGCGGGAGCCGCCCGAGGCAGCUUCCCCGUGACAGAGCCUCAGCCUGCCUGGAAG
          -500                                                  -450                                                  -400

                                                -120
Met Pro Arg Leu Cys Ser Ser Arg Ser Gly Ala Leu Leu Leu Ala Leu Leu Leu Gln Ala Ser Met Glu Val Arg Gly Trp Cys Leu Glu
AUG CCG AGA CUG UGC AGC AGU CGU UCG GGC GCC CUG CUG CUG GCC UUG CUG CUU CAG GCC UCC AUG GAA GUG CGU GGU UGG UGC CUG GAG
                                                        -350
   -100                                                                                 -80
Ser Ser Gln Cys Gln Asp Leu Thr Thr Glu Ser Asn Leu Leu Ala Cys Ile Arg Ala Cys Lys Pro Asp Leu Ser Ala Glu Thr Pro Val
AGC AGC CAG UGU CAG GAC CUC ACC ACG GAA AGU AAC CUG CUG GCG UGC AUC CGG GCC UGC AAG CCC GAC CUC UCC GCC GAG ACG CCG GUG
  -300                                                                  -250
                                                -60
Phe Pro Gly Asn Gly Asp Glu Gln Pro Leu Thr Glu Asn Pro Arg Lys Tyr Val Met Gly His Phe Arg Trp Asp Arg Phe Gly Arg Arg
UUC CCC GGC AAC GGC GAU GAG CAG CCG CUG ACU GAG AAC CCC CGG AAG UAC GUC AUG GGC CAU UUC CGC UGG GAC CGC UUC GGC CGU CGG
               -200                                                                 -150
   -40                                                                                  -20
Asn Gly Ser Ser Ser Ser Gly Val Gly Gly Ala Ala Gln Lys Arg Glu Glu Glu Val Ala Val Gly Glu Gly Pro Gly Pro Arg Gly Asp
AAU GGU AGC AGC AGC AGC GGA GUU GGG GGC GCG GCC CAG AAG CGC GAG GAG GAA GUG GCG GUG GGC GAA GGC CCC GGG CCC CGC GGC GAU
                           -100                                                                 -50
                                          -1   1
Asp Ala Glu Thr Gly Pro Arg Glu Asp Lys Arg Ser Tyr Ser Met Glu His Phe Arg Trp Gly Lys Pro Val Gly Lys Lys Arg Arg Pro
GAC GCC GAG ACG GGU CCG CGC GAG GAC AAG CGU UCU UAC UCC AUG GAA CAC UUC CGC UGG GGC AAG CCG GUG GGC AAG AAG CGG CGC CCG
                                        -1 1                                                                    50
 20                                                                                 40
Val Lys Val Tyr Pro Asn Gly Ala Glu Asp Glu Ser Ala Gln Ala Phe Pro Leu Glu Phe Lys Arg Glu Leu Thr Gly Glu Arg Leu Glu
GUG AAG GUG UAC CCC AAC GGC GCC GAG GAC GAG UCG GCC CAG GCC UUU CCC CUC GAA UUC AAG AGG GAG CUG ACC GGG GAG AGG CUC GAG
                                                    100
                                          60
Gln Ala Arg Gly Pro Glu Ala Gln Ala Glu Ser Ala Ala Ala Arg Ala Glu Leu Glu Tyr Gly Leu Val Ala Glu Ala Glu Ala Glu Ala
CAG GCG CGC GGC CCC GAG GCC CAG GCU GAG AGU GCG GCC GCC CGG GCU GAG CUG GAG UAU GGC CUG GUG GCG GAG GCG GAG GCU GAG GCG
 150                                                                  200
 80                                                                                 100
Ala Glu Lys Lys Asp Ser Gly Pro Tyr Lys Met Glu His Phe Arg Trp Gly Ser Pro Pro Lys Asp Lys Arg Tyr Gly Gly Phe Met Thr
GCC GAG AAG AAG GAC UCG GGG CCC UAU AAG AUG GAA CAC UUC CGC UGG GGC AGC CCG CCC AAG GAC AAG CGC UAC GGC GGG UUC AUG ACC
               250                                                                  300
                                          120
Ser Glu Lys Ser Gln Thr Pro Leu Val Thr Leu Phe Lys Asn Ala Ile Ile Lys Asn Ala His Lys Lys Gly Gln
UCC GAG AAG AGC CAA ACG CCC CUU GUC ACG CUG UUC AAA AAC GCC AUC AUC AAG AAC GCC CAC AAG AAG GGC CAG UGA GGGCGCAGCGGGCAG
                           350                                                                  400

GGGCCUCUCUCCGCGGAAAGUUGACCCUGAAGGCCUCUCUUCUGCCCUCCUACCGCCUCGCAGCCUGGGUGAGGAUUCGCCCAGGCAGUGAUGGCGCCAGGUAUCCCGACUCUUAAAGC
                           450                                                                  500

UGUCUGUAGUUAAGAAAUAAAACCUUUCAAGUUUCACGAAAAAAAA----3'
          550
```

Fig. 2 Primary structure of bovine ACTH–β-LPH precursor mRNA. The nucleotide sequence of the ACTH–β-LPH precursor mRNA was deduced from that of the cDNA insert in pSNAC20. The nucleotide numbers (see legend to Fig. 1) are indicated below the sequence. The predicted amino acid sequence up to the methionine residue encoded by nucleotide residues −393 to −391 is shown above the nucleotide sequence. Amino acid residues are numbered in the direction from amino terminus to carboxyl terminus, beginning with the first residue of ACTH, and the residues on the amino-terminal side of ACTH are indicated by negative numbers; the amino acid numbers are given above the sequence. The sequence of nucleotide residues 1–207 was reported previously[17] and was confirmed by sequencing additional restriction fragments as shown in Fig. 1. The cDNA insert includes a poly(A) tract of eight nucleotides at positions 578–585, but the average length of the poly(A) sequence in the mRNA was previously estimated to be 68 nucleotides[4].

©1979 Nature Publishing Group

methionine and tryptophan residues at the positions characteristic of the MSH structure and shares with α-MSH serine and glutamic acid residues at equivalent positions. Computer analysis indicates no structural correspondence of other amino acid sequences in the cryptic portion of the precursor with hormones of known sequence[39].

From the findings described above, we conclude that the ACTH–β-LPH precursor protein is probably composed of four repetitive structures as shown in Fig. 4. Each unit contains a homologous peptide core segment having an MSH sequence or its analogue, which is followed by or lies within an individually different sequence. Furthermore, each of the repetitive units, as well as the MSH sequences themselves, is flanked on both sides by a pair of basic amino acid residues (or by one of the termini of the precursor molecule). The putative structure of the common precursor protein suggests that each of its units is destined to be separated from its neighbour by proteolytic processing and that further cleavage occurs within the unit to yield smaller peptides.

The peptide fragment consisting of amino acid residues −131 to −58 includes five cysteine residues, and the remainder of the precursor molecule contains no cysteine residue. Other peptide hormones, such as growth hormone and insulin, contain an even number of cysteine residues which form disulphide bonds[39]. Thus, we can assume that the cysteine residue at position −127 is a constituent of the putative signal peptide, which is removed during the process of secretion, although we cannot exclude the possibility that an odd number of cysteine residues are encoded by the missing portion of the mRNA.

The repetitive nature of the amino acid sequence of the ACTH–β-LPH precursor molecule raises interesting questions concerning both the biological significance of its characteristic structure and the biological functions of possible peptide fragments that may be derived from it. The presence of several partly homologous units within the same molecule suggests that the functions of the component peptides may be related to or coordinated with one another. It is well established that ACTH is essential for the induction and maintenance of glucocorticoid production in adrenocortical cells[40]. However, additional pituitary effects on adrenocortical cells, such as stimulation of mitogenic activity and mineralocorticoid production, cannot be accounted for solely by the action of ACTH[40]. It is an intriguing hypothesis that other peptides which may be elaborated from the precursor molecule are involved in these actions. Another possibility is that these structurally related peptides are involved in neural functions. Several studies have demonstrated that not only endorphins and methionine-enkephalin[41] but also MSHs and related peptides[42,43] are present in various regions of the central nervous system. In addition to the potent opiate activity of endorphins and methionine-enkephalin[44], MSH-related peptides have been reported to exert certain effects on brain function, such as learning and retention of new behaviour[45]. Thus, it is attractive to speculate that various peptides derived from the ACTH–β-LPH precursor may act together or antagonistically as modulators of neural functions.

Table 1 Codon usage in bovine ACTH–β-LPH precursor mRNA

Phe	UUU	1	Ser	UCU	1	Tyr	UAU	2	Cys	UGU	1
	UUC	8		UCC	4		UAC	4		UGC	4
Leu	UUA	0		UCA	0	Ter	UAA	0	Ter	UGA	1
	UUG	1		UCG	3		UAG	0	Trp	UGG	4
Leu	CUU	2	Pro	CCU	0	His	CAU	1	Arg	CGU	4
	CUC	4		CCC	11		CAC	3		CGC	10
	CUA	0		CCA	0	Gln	CAA	1		CGA	0
	CUG	13		CCG	7		CAG	9		CGG	5
Ile	AUU	0	Thr	ACU	1	Asn	AAU	1	Ser	AGU	3
	AUC	3		ACC	3		AAC	6		AGC	9
	AUA	0		ACA	0	Lys	AAA	1	Arg	AGA	1
Met	AUG	6		ACG	5		AAG	18		AGG	2
Val	GUU	1	Ala	GCU	3	Asp	GAU	2	Gly	GGU	3
	GUC	2		GCC	16		GAC	8		GGC	17
	GUA	0		GCA	0	Glu	GAA	7		GGA	1
	GUG	8		GCG	8		GAG	22		GGG	5

The amino acid sequence of the bovine ACTH–β-LPH precursor has been predicted solely from the nucleotide sequence of the mRNA, assuming that the translational initiation site is located at the methionine residue at position −131. Numbers next to codons indicate the numbers of amino acids using particular codons.

Characteristic features of ACTH–β-LPH precursor mRNA

We have previously observed that the coding regions for ACTH and the first segment of β-LPH contain an unusually large number of repeated nucleotide sequences, suggesting that these regions of the structural gene for the ACTH–β-LPH precursor have evolved by a series of genetic duplications[17]. In the present study, this observation has been extended to the remaining portions of the ACTH–β-LPH precursor mRNA. Comparison of the coding regions for α-MSH, β-MSH and γ-MSH strongly supports the view that DNA sequence duplications have occurred during the evolution of these genes (Fig. 3). The nucleotide sequences encoding the common amino acids are conserved, except for a very few silent codon changes; however, those coding for the amino acids that are different in the three peptides are drastically altered. In addition to showing the extensive nucleotide sequence homology just mentioned, computer analysis[46,47] of the nucleotide sequence data presented in Fig. 2 reveals that the ACTH–β-LPH precursor mRNA contains other segments that seem to represent direct duplications; for example, residues −453 to −443 and 406 to 417; −93 to −82 and 235 to 246; −65 to −54 and 35 to 45; −11 to −2 and 299 to 308; 163 to 187 and 220 to 241; 400 to 411 and 506 to 517. Taken together, these findings support our concept that the structural gene for the ACTH–β-LPH precursor protein has evolved by a series of direct duplications of certain common ancestral DNA segments followed by substitution, addition or deletion of some regions.

Codon choices for the ACTH–β-LPH precursor mRNA have been assigned (Table 1), using the amino acid sequence shown in Fig. 2. No definitive conclusion can be reached concerning codon utilisation because the initiation site for translation has not been established. Nevertheless, codon usage seems to occur nonrandomly, as in the case of other eukaryotic mRNAs[21–25], and specific codon choices seem to be strongly preferred for

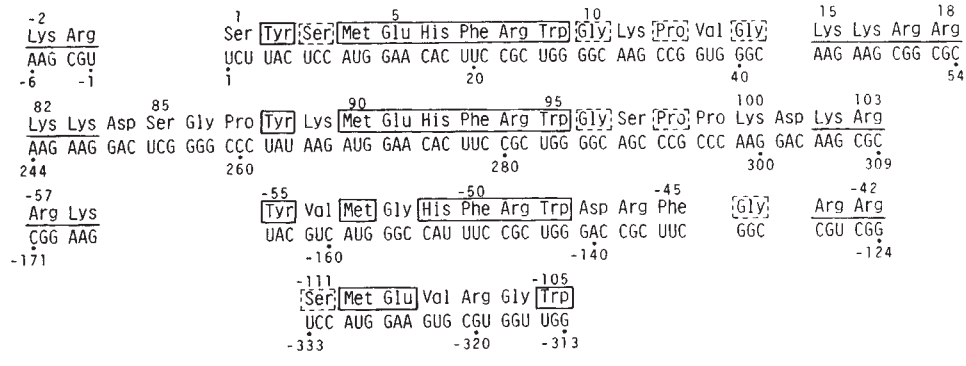

Fig. 3 Comparison of nucleotide and amino acid sequences in the regions including MSHs and their analogue. The amino acid sequences are lined up, introducing gaps so as to point out the homologies discussed in the text; for the numbering of nucleotide and amino acid residues, see legends to Figs 1 and 2. The amino acid residues in solid-line boxes occur in at least three of the four structures shown and those in dashed-line boxes are found in two of them. Sequences of two or four consecutive basic amino acid residues are underlined. α-MSH[39] represents amino acid residues 1–13, of which the serine residue at position 1 is N-acetylated and the valine residue at position 13 is in the form of an amide; β-MSH[39] represents amino acid residues 84–101.

©1979 Nature Publishing Group

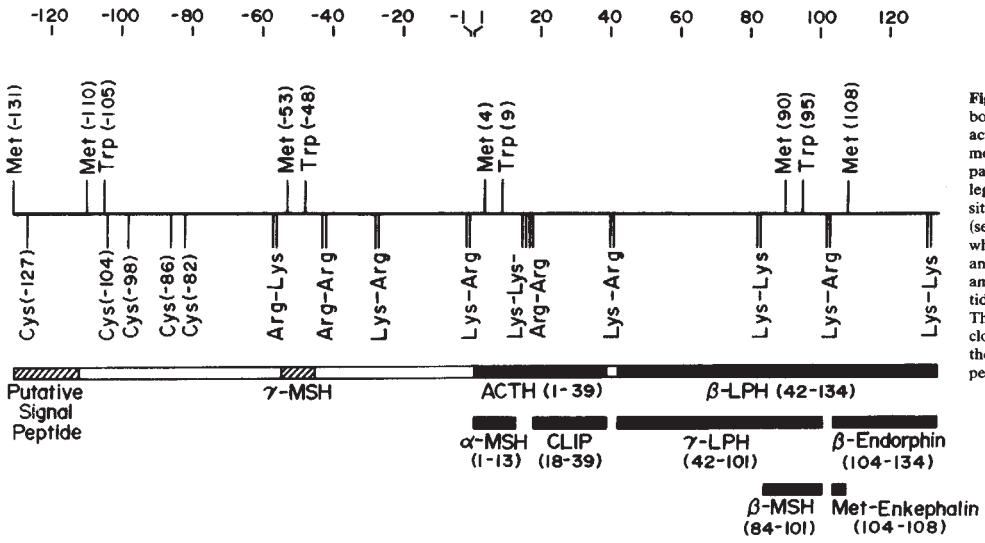

Fig. 4 Schematic representation of the structure of bovine ACTH–β-LPH precursor. Characteristic amino acid residues are shown, and the positions of the methionine, tryptophan and cysteine residues are given in parentheses; for the numbering of amino acid residues, see legend to Fig. 2. The location of the translational initiation site at the methionine residue at position −131 is assumed (see the text). The closed bars represent the regions for which the amino acid sequence was known, and the open and the shaded bars represent the regions for which the amino acid sequence has been predicted from the nucleotide sequence of the ACTH–β-LPH precursor mRNA. The locations of known component peptides are shown by closed bars; the amino acid numbers are given in parentheses. The locations of γ-MSH and the putative signal peptide are indicated by shaded bars; the termini of these peptides are not definitive.

some amino acids. Certain codons (for example, UUC for Phe, CUG for Leu, AAG for Lys) are used repeatedly, whereas other codon choices are either not made or are rare (for example, UUA for Leu, AUA for Ile, GUA for Val). This mode of codon selection reflects an apparent preference for G or C of the third position of mammalian codons[21–24]. Assuming that the initiation site is located at the methionine residue at position −131, 116 codons (44%) terminate in G, 112 (42%) in C, 26 (10%) in U and 11 (4%) in A.

Reflecting the selective codon usage, the ACTH–β-LPH precursor mRNA (excluding the poly(A) region) has a high G+C content (65%) which exceeds that of total bovine DNA (39%)[48]. Also, this mRNA exhibits nonrandom dinucleotide frequencies. The abundance of the dinucleotide CG in this mRNA exhibits a marked contrast to its deficiency in other eukaryotic mRNAs[21–25] as well as in total bovine DNA[49].

The 3′ noncoding region of the ACTH–β-LPH precursor mRNA also shares some features with other eukaryotic mRNAs[21–25,50]. A lengthy inverted repeat (that is, region of two-fold rotational symmetry) exists at nucleotide residues 469–492 and 521–497, and this inverted repeat follows two in-frame copies of the nonsense codon UGA. Similar sequence features have been reported for the 3′ noncoding end of human chorionic somatomammotropin mRNA[51]. Also, the sequence AAUAAA (residues 555–560), which is commonly present near the 3′ noncoding end of eukaryotic mRNAs, is found in the corresponding region of the ACTH–β-LPH precursor mRNA.

The sequencing data are available from the authors on request.

We thank Drs M. Takanami and K. Sugimoto for their advice and for gifts of some restriction endonucleases, and Dr T. Honjo for advice. The investigations carried out in Kyoto were supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation, the Foundation for the Promotion of Research on Medicinal Resources, and the Japanese Foundation of Metabolism and Diseases. The studies at Stanford were supported by grants to S.N.C. from the NIH, the American Cancer Society, and the NSF.

Received 29 January; accepted 12 February 1979.

1. Nakanishi, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **73,** 4319–4323 (1976).
2. Nakanishi, S., Inoue, A., Taii, S. & Numa, S. *FEBS Lett.* **84,** 105–109 (1977).
3. Taii, S., Nakanishi, S. & Numa, S. *Eur. J. Biochem.* **93,** 205–212 (1979).
4. Kita, T., Inoue, A., Nakanishi, S. & Numa, S. *Eur. J. Biochem.* **93,** 213–220 (1979).
5. Mains, R. E. & Eipper, B. A. *J. biol. Chem.* **251,** 4115–4120 (1976).
6. Mains, R. E., Eipper, B. A. & Ling, N. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3014–3018 (1977).
7. Eipper, B. A. & Mains, R. E. *J. biol. Chem.* **253,** 5732–5744 (1978).
8. Roberts, J. L. & Herbert, E. *Proc. natn. Acad. Sci. U.S.A.* **74,** 4826–4830, 5300–5304 (1977).
9. Scott, A. P. *et al. Nature new Biol.* **244,** 65–67 (1973).
10. Li, C. H. & Chung, D. *Proc. natn. Acad. Sci. U.S.A.* **73,** 1145–1148 (1976).
11. Ling, N., Burgus, R. & Guillemin, R. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3942–3946 (1976).
12. Nakanishi, S., Kita, T., Taii, S., Imura, H. & Numa, S. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3283–3286 (1977).
13. Nakamura, M., Nakanishi, S., Sueoka, S., Imura, H. & Numa, S. *Eur. J. Biochem.* **86,** 61–66 (1978).
14. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. *Proc. natn. Acad. Sci. U.S.A.* **70,** 3240–3244 (1973).
15. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
16. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87,** 107–110 (1978).
17. Nakanishi, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 6021–6025 (1978).
18. Timmis, K. N., Cabello, F. & Cohen, S. N. *Molec. gen. Genet.* **167,** 11–19 (1978).
19. Chang, A. C. Y., Lansman, R. A., Clayton, D. A. & Cohen, S. N. *Cell* **6,** 231–244 (1975).
20. Brutlag, D., Fry, K., Nelson, T. & Hung, P. *Cell* **10,** 509–519 (1977).
21. Efstratiadis, A., Kaftos, F. C. & Maniatis, T. *Cell* **10,** 571–585 (1977).
22. Ullrich, A. *et al. Science* **196,** 1313–1319 (1977).
23. Seeberg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270,** 486–494 (1977).
24. Shine, J., Seeburg, P. H. Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270,** 494–499 (1977).
25. McReynolds, L. *et al. Nature* **273,** 723–728 (1978).
26. Chang, A. C. Y. *et al. Nature* **275,** 617–624 (1978).
27. Li, C. H., Tan, L. & Chung, D. *Biochem. biophys. Res. Commun.* **77,** 1088–1093 (1977).
28. Pankov, Y. A. *Vop. med. Khim.* **19,** 330–332 (1973).
29. Steiner, D. F., Kemmler, W., Tager, H. S. & Peterson, J. D. *Fedn Proc.* **33,** 2105–2115 (1974).
30. Habener, J. F., Potts, J. T. Jr & Rich, A. *J. biol. Chem.* **251,** 3893–3899 (1976).
31. Habener, J. F., Chang, H. T. & Potts, J. T. Jr *Biochemistry* **16,** 3910–3917 (1977).
32. Weber, K. & Osborn, M. in *The Proteins* Vol. 1, 3rd edn (eds Neurath, H. & Hill, R. L.) 179–223 (Academic, New York, 1975).
33. Devillers-Thiery, A., Kindt, T., Scheele, G. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **72,** 5016–5020 (1975).
34. Chan, S. J., Keim, P. & Steiner, D. F. *Proc. natn. Acad. Sci. U.S.A.* **73,** 1964–1968 (1976).
35. Kemper, B., Habener, J. F., Ernst, M. D., Potts, J. T. Jr & Rich, A. *Biochemistry* **15,** 15–19 (1976).
36. Lingappa, V. R., Devillers-Thiery, A. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **74,** 2432–2436 (1977).
37. Lowry, P. J., Hope, J. & Silman, R. E. *Excerpta med. int. Cong. Ser.* **402,** 71–76 (1977).
38. Schwyzer, R. *Ann. N.Y. Acad. Sci.* **297,** 3–26 (1977).
39. Dayhoff, M. D. in *Atlas of Protein Sequence and Structure* Vol. 5, 173–227 (National Biomedical Research Foundation, Washington, D.C., 1972).
40. Urquhart, J. in *Handbook of Physiology* sect. 7, Vol. 4, part 2 (eds Greep, R. O. & Astwood, E. B.) 133–157 (American Physiological Society, Washington, D.C., 1974).
41. Rossier, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **74,** 5162–5165 (1977).
42. Krieger, D. T., Liotta, A. & Brownstein, M. J. *Proc. natn. Acad. Sci. U.S.A.* **74,** 648–652 (1977).
43. Oliver, C. & Proter, J. C. *Endocrinology* **102,** 697–705 (1978).
44. Goldstein, A. *Science* **193,** 1081–1086 (1976).
45. Reith, M. E. A., Schotman, P., Gispen, W. H. & de Wied, D. *Trends biochem. Sci.* **2,** 56–58 (1977).
46. Korn, L. J., Queen, C. L. & Wegman, M. N. *Proc. natn. Acad. Sci. U.S.A.* **74,** 4401–4405 (1977).
47. Staden, R. *Nucleic Acids Res.* **4,** 4037–4051 (1977).
48. Szybalski, W. *Meth. Enzym.* **12B,** 330–360 (1968).
49. Swartz, M. N., Trautner, T. A. & Kornberg, A. *J. biol. Chem.* **237,** 1961–1967 (1962).
50. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263,** 211–214 (1976).
51. Seeburg, P. H. *et al. Cell* **12,** 157–165 (1977).
52. Sutcliffe, J. G. *Nucleic Acids Res.* **5,** 2721–2728 (1978).
53. Hayward, G. S. *Virology* **49,** 342–344 (1972).

©1979 Nature Publishing Group