Postprint version. Original publication in:
Molecular Biology (USSR) (1988) 22, 1, 117 - 121
Moleculyarnaya Biologiya (1988) 22, 1, 139 - 144

# GENETIC ENGINEERING OF PEPTIDE HORMONES.

## II. POSSIBLE POLYMORPHISM OF PREPROLACTIN IN CATTLE. DATA OF MOLECULAR CLONING

P. M. Rubtsov, R. G. Oganesyan, V. G. Gorbulev,
K. G. Skryabin, and A. A. Baev

UDC 577.212.175.3

Primary structure is determined of an insertion of a clone isolated from the library of hypophyseal cDNA of cattle by hybridization with a probe specific for prolactin. Analysis of nucleotide sequences showed that in the process of cloning, reorganization occurred in structure of preprolactin cDNA, including an inversion of the 5'-terminal and deletion of the central section of cDNA. Nevertheless, from structure of cDNA, nucleotide sequences can be deduced of extended 5'- and 3'-terminal sections of preprolactin mRNA in cattle with lengths of 257 and 551 nucleotide residues, respectively. When these sequences are compared to those established previously, some differences were found in primary structure. The most important of them is the presence of an additional codon which codes alanine at the position (–22) of the signal peptide. It is suggested that heterogeneity of preprolactin mRNA of cattle in the section coding the signal peptide is the result of alternative splicing, as was shown for preprolactin mRNA in rats.

Prolactin is a peptide hormone of the hypophysis. In the same way as related hormones (growth hormone, placental lactogen), it is synthesized in the form of a precursor, preprolactin, which differs from the mature hormone by the presence of an N-terminal signal peptide. Synthesis and cloning of cDNA of prolactins of the rat [1–3], cattle [4], and man [5] permitted the structure of precursors of these hormones to be refined [1-5]. When determining nucleotide sequences of cDNA of rat preprolactin, two types of sequences were identified which differ in structure of the section coding the signal peptide. Two groups of authors have published nucleotide sequences which contained a GCA triplet which codes alanine at the position (–20) of the signal peptide [1, 2]. A third sequence [3] did not contain this triplet and coded preprolactin, which has a signal peptide one amino acid residue shorter. Determination of nucleotide sequence of the gene for preprolactin in the rat showed that the splicing section at the border of intron A and exon 2 has a structure which allows the possibility of alternative splicing [6, 7]. The suggestion was made that heterogeneity of preprolactin mRNA in the rat is a result of such alternative splicing during excision of intron A. Nucleotide sequences were later determined for preprolactin cDNA of cattle [4] and man [5]. In both cases the sequences coding the signal peptide of these hormones had large homology to the analogous sequence of preprolactin cDNA in the rat, but which did not contain the alanine triplet at the position corresponding to the codon (–20) of the signal peptide of rat preprolactin. It was unclear whether the alternative splicing of preprolactin pre-mRNA is a unique feature of the rat hypophysis or whether an analogous phenomenon can take place during processing of pre-mRNA of preprolactins of other mammals.

We previously reported cloning of cDNA of cattle preprolactin [8]. Results are presented in the present work of determining the nucleotide sequence of a cDNA-insertion of one of the clones obtained. It follows from results of sequencing that reorganization of the structure of cDNA occurred in the process of cloning. At the same time, the nucleotide sequence of extended 5'- and 3'-terminal sections of preprolactin mRNA of cattle can be deduced from the structure of cDNA. Comparison of this sequence with that published previously [4] permits the conclusion to be made that heterogeneity of the signal peptide of preprolactin is also characteristic of cattle.

Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow. Translated from Molekulyarnaya Biologiya, Vol. 22, No. 1, pp. 139-144, January-February, 1988.

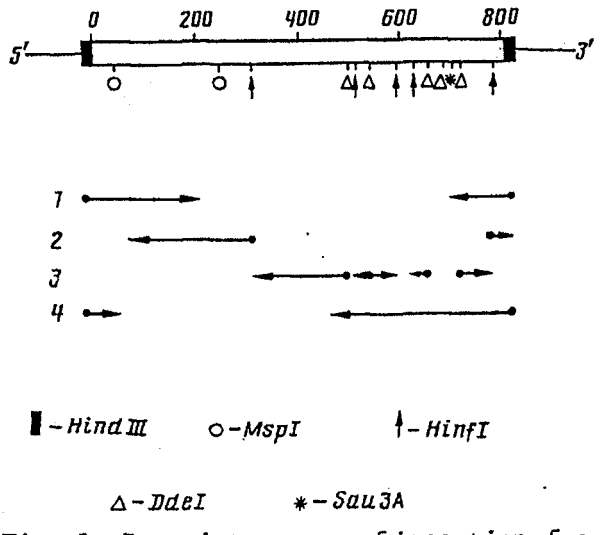


Fig. 1. Restrictase map of insertion from plasmid pBPL6 and strategy of sequencing it.

## EXPERIMENTAL

Restrictases Sau3A, HindIII, EcoRI, BamHI, and MspI obtained from the Ferment nongovernmental organization were used in the work, along with HinfI and DdeI from P–L Biochemicals, Klenow fragment of DNA-polymerase I from Bethesda Research Laboratories, and [$\alpha^{-32}$]dNTP was obtained from Amersham and the Izotop All-Union Association.

Isolation of plasmid DNA, cleavage by restrictases, electrophoresis in agarose and polyacrylamide gels, isolation and purification of DNA fragments, and insertion of label into 3'-ends of fragments were performed according to recommendations presented in the work of Maniatis, et al. [9]. Nucleotide sequence of DNA was determined according to Maxam and Gilbert [10].

## RESULTS AND DISCUSSION

Analysis of plasmids isolated by hybridization of colonies showed that all of them contained an insertion of cDNA more than 500 basepairs (b.p.) in size. For more detailed investigation, a plasmid designated as pBPL6 and carrying an insertion ∿800 b.p. in size was chosen. The restrictase map of the insertion from plasmid pBPL6 and the strategy of sequencing are shown in Fig. 1. It follows from determination of the complete nucleotide sequence of the insertion that its length constitutes 808 b.p., not counting sequences of the HindIII linker (Fig. 2). Comparison of the nucleotide sequence determined by us with that published previously [4] showed that pBPL6 contains a reorganized sequence of cattle preprolactin cDNA. The character of the reorganizations are presented schematically in Fig. 3. It is seen that the insertion contains a continuous sequence from the codon of the 67th amino acid of prolactin to the terminal codon and an untranslated section on the 3'-end 146 b.p. long. The nucleotide sequence coding the 39th through 65th amino acid of prolactin is completely absent in cloned cDNA. At the same time, an untranslated sequence of the 5'-end (50 b.p.) and the sequence coding the signal peptide and first 38 amino acids of the mature hormone are represented, but in inverted form in relation to the 3'-terminal section. Reorganizations of this type, which occur during synthesis and cloning of cDNA, have been described in the literature, for example for cDNA of influenza virus [11], human interferon [12], bovine parathyroid hormone [13], and pigeon α-globulin [14].

On the whole there are two sections in the cDNA cloned by us which constitute a total of more than 90% of full-size cDNA of cattle proprolactin. Sequences corresponding to them of 5'- and 3'-terminal sections of preprolactin mRNA measuring 257 and 551 nucleotides, respectively, are presented in Fig. 4.

Comparison of these sequences with those published previously [4] revealed several differences in primary structure. The most important is the presence in the sequence presented by us of an additional GCA triplet which codes alanine at the position (–22) of the signal peptide (Fig. 5). Considering this amino acid, the length of the signal peptide of cattle preprolactin constitutes 31 amino acid residues. This result does not agree with data pre-

```
TTGAACATTT CCGAGGAGAG GTCATGCATG TAGTGGCACA GCATGACTGC CCGGTCAAAG  60
AGGTCTCGAA GGGATACCTG GCAGTTGCCA GGCCCATTGG GACAGACGGG GGTGGAGACC 120
ACACCCTGGC ACAAGAGTAG ATTTGACACC AGCAGCAGCA GGAGCAGGCG GGACCCTGCT 180
TTCTGCGACG AACCTTTGCT GTCCATGGTG GTGATGATTT CAAGAAACAC ACTTCACCAG 240
GAAGCTCTCG TCCTATCCCC GAAGATAAAG AACAAGCCCA ACAGACCCAT CATCAAGTCC 300
TTATGAGCTT GATTCTTGGG TTGCTGCCCT CCTGGAATGA CCCTCTCTAT CACCTAGTCA 360
CCGAGGTGCC GGGTATGAAA GGAGCCCCAG ATGCTATCCT ATCGACGGCC ATAGAGATTC 420
AGGAAGAAAA CAAACGACTT CTGGAAGGCA TGGAGATGAT ATTTGGCCAG GTTATTCCTG 480
GAGCCAAAGA GACTGAGCCC TACCCTGTGT GGTCAGGACT CCCGTCCCTG CAAACTAAGG 540
ATGAAGATGC ACGTTATTCT GCTTTTTATA ACCTGCTCCA CTGCCTGGCC AGGGATTCAA 600
GCAAGATTGA CACTTACCTT AAGCTCCTGA ATTGCAGAAT CATCTACAAC AACAACTGCT 660
AAGCCCACAT TCCATCCTAT CCATTTCTGA GATGGTTCTT AATGATCCAT TCCCTGGCAA 720
ACTTCTCTGA GCTTTATAGC TTTGTAATGC ATGCTTGGCT CTAATGGGTT TCATCTTAAA 780
TAAAAACAGA CTCTGTAGCG ATGTCAAA
```

Fig. 2. Sequence of nucleotides of insertion from plasmid pBPL6.

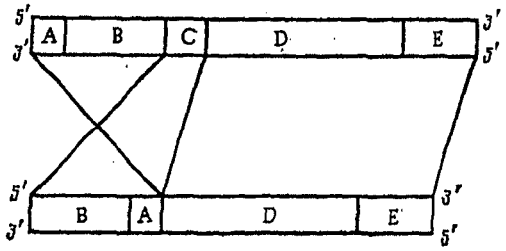


Fig. 3. Diagram of reorganization of nucleotide sequence of cDNA of cattle preprolactin in plasmid pBPL6: A) 5'-untranslated sequence; B) sequences coding signal peptide and first 38 amino acids of prolactin; C) sequence coding amino acids 39-65; D) sequence coding amino acids 66-199; E) 3'-untranslated sequence.

sented in other works [4, 15], in which the size of the signal peptide was estimated to be 30 amino acid residues. The cDNA cloned by us probably represents a copy of a minor mRNA fraction. As in the case of rat preprolactin mRNA, heterogeneity in the section coding the signal peptide is apparently the result of alternative splicing. At the present time the nucleotide sequence has been established for the prolatin genes of the rat [6, 7], man [16], and cattle [17] while the structure of the exon-intron transition in these genes rules out the possibility of an alternative case. In the case of rat preprolactin mRNA, the possibility of ambiguous excision of intron A was demonstrated by cloning two variants of cNDA [1, 3]. Direct determination of the nucleotide sequence of mRNA, conducted by Taylor et al. [2], indicates that the basic variant is the one containing the alanine codon. In the case of cattle preprolactin mRNA, the variant not containing this codon is probably more richly represented. The only published structure of human preprolactin cDNA [5] also does not contain an alanine codon in this region.

Whether the selection of a section in which intron A is excised is a random event or is determined by some kind of actor, hormonal status of the organism, for example, remains unknown.

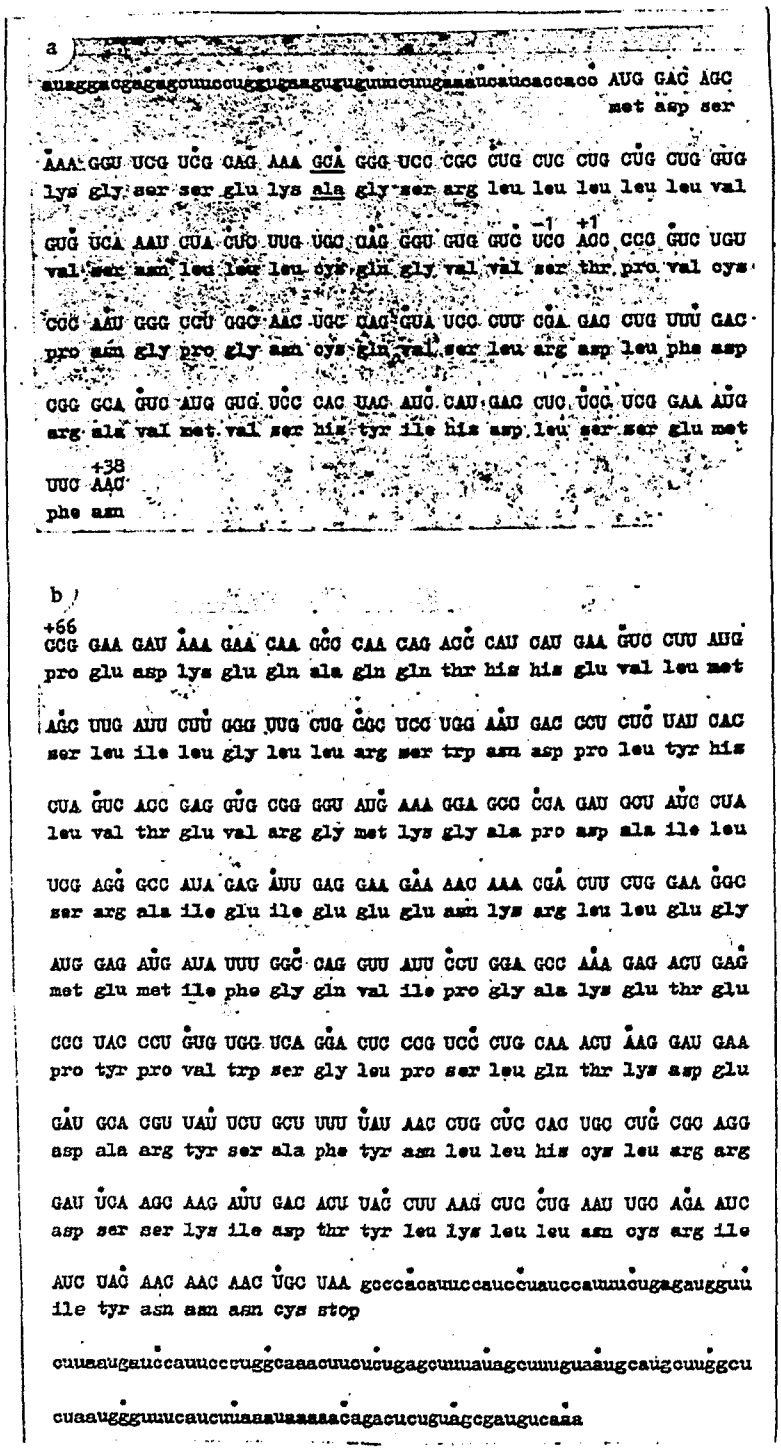


Fig. 4. Autoradiogram of structural gel and nucleotide sequence read from it which codes the signal peptide section of cattle preprolactin.

Attention should be turned to the fact that alternative splicing has also been described for pre-mRNA of growth hormone related to prolactin. In the gene for human growth hormone, intron B has two alternative sections of splicing at the 3'-end. When the basic section functions, normal growth hormone forms; splicing at the other section leads to formation of the so-called 20 kilodalton variant of growth hormone, which has a deletion of the amino acid sequence in the section from the 32nd through 46th amino acid residue [18].

In conclusion, it should be noted that differences in structure of signal peptides of two variants of cattle preprolactin should not be reflected in their function, since an addi-

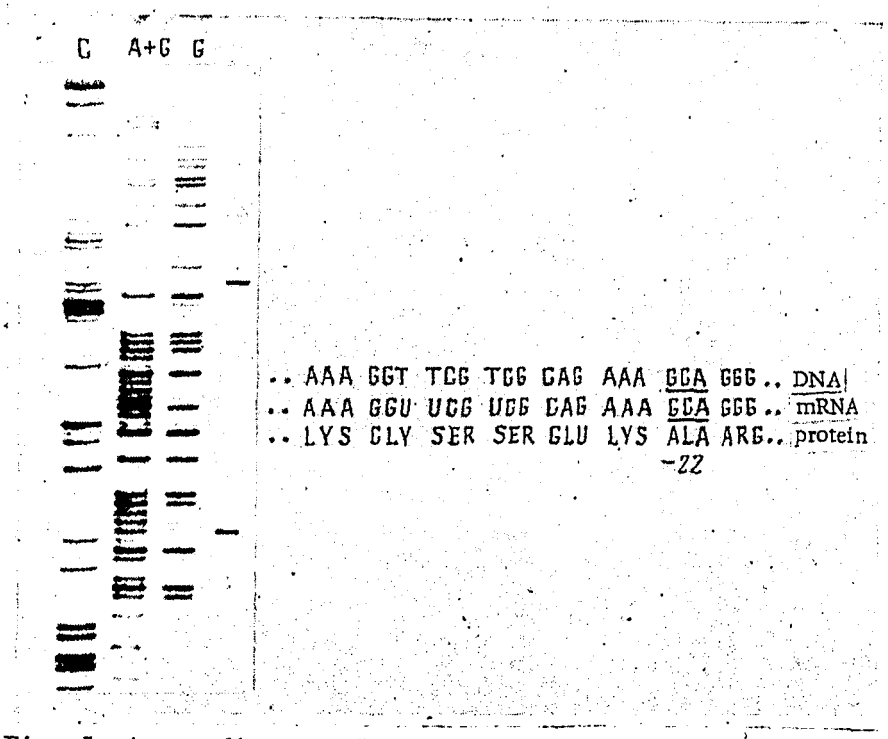


Fig. 5. Autoradiogram of structural gel and nucleotide sequence read from it which codes the signal peptide section of cattle preprolactin.

tional alanine residue is found in the section in front of the extended hydrophobic sequence and the overall topography of the signal peptide is not disrupted.

## LITERATURE CITED


1. E. J. Gubbins, R. A. Maurer, M. Lagrimini, C. R. Erwin, and J. E. Donelson, J. Biol. Chem. 255, 8655-8662 (1980).
2. W. L. Taylor, K. J. Collier, H. Lee Weith, and J. E. Dixon, Biochem. Biophys. Res. Commun., 102, 1071-1077 (1981).
3. N. E. Cooke, D. Coit, R. I. Weiner, J. D. Baxter, and J. A. Martial, J. Biol. Chem., 255, 6502-6510 (1980).
4. N. L. Sasavage, J. H. Nilson, S. Horowitz, and F. M. Rottman, J. Biol. Chem., 257, 678-681 (1982).
5. N. E. Cooke, D. Coit, J. Shine, J. D. Baxter, and J. A. Martial, J. Biol. Chem., 256, 4007-4016 (1981).
6. R. A. Maurer, C. R. Erwin, and J. E. Donelson, J. Biol. Chem., 256, 10524-10528 (1981).
7. N. E. Cooke and J. D. Baxter, Nature, 297, 603-606 (1982).
8. P. M. Rubtsov, B. K. Chernov, V. G. Gorbulev, A. Sh. Parsadanyan, P. S. Sverdlova, V. V. Chupeeva, Yu. B. Golova, N. B. Batchikova, G. S. Zhvirblis, K. G. Skryabin, and A. A. Baev, Mol. Biol., 19, 267-277 (1985).
9. T. Maniatis, E. F. Fritsch, and J. Sambrook, Molecular Cloning, Cold Spring Harbor Lab., Cold Spring Harbor, New York (1982).
10. A. M. Maxam and W. Gilbert, Proc. Natl. Acad. Sci. USA, 74, 560-564 (1977).
11. S. Fields and G. Winter, Gene, 15, 207-214 (1981).
12. G. Volskaert, J. Tavernier, R. Derynck, R. Devos, and W. Fiers, Gene, 15, 215-223 (1981).
13. C. A. Weaver, D. F. Gordon, and B. Kemper, Proc. Natl. Acad. Sci. USA, 78, 4073-4077 (1981)
14. A. G. Pletnev, N. A. Scobeleva, L. Yu. Frolova, and L. L. Kisselev, Gene, 25, 93-100 (1983).
15. V. R. Lingappa, A. Devillers-Thiery, and G. Blobel, Proc. Natl. Acad. Sci. USA, 74, 2432-2436 (1977).
16. A. T. Truong, C. Duez, A. Belayev, A. Renard, R. Pictet, G. I. Bell, and J. A. Martial, EMBO J., 3, 429-437 (1984).
17. S. A. Camper, D. N. Luck, Y. Yao, R. P. Woychik, R. G. Goodwin, R. N. Lyons, Jr., and F. M. Rottman, DNA, 3, 237-249 (1984).
18. A. T. De Noto, D. D. Moore, and H. M. Goodman, Nucleic Acids Res., 9, 3719-3730 (1981).