Complete nucleotide sequence of the human corticotropin-β-lipotropin precursor gene

Hideo Takahashi, Yasuhiro Hakamata, Yumiko Watanabe, Reiko Kikuno*, Takashi Miyata* and Shosaku Numa

Department of Medical Chemistry, Kyoto University Faculty of Medicine, Kyoto 606, and *Department of Biology, Kyushu University Faculty of Science, Fukuoka 812, Japan

Received 16 August 1983; Accepted 7 September 1983

ABSTRACT

The nucleotide sequence of an 8658-base-pair human genomic DNA segment containing the entire corticotropin-β-lipotropin precursor gene has been determined, and some sequence features of the gene and its flanking regions have been analysed. The gene is composed of 7665 base pairs including two introns of 3708 and 2886 base pairs. Comparison of the 5'-flanking sequences of the human, bovine and mouse corticotropin-β-lipotropin precursor genes reveals the presence of a highly conserved region, which contains sequences of 14-15 base pairs homologous with sequences located upstream of the mRNA start site of other glucocorticoid-regulated genes.

INTRODUCTION

The primary structure of the common precursor of the pituitary hormones corticotropin (ACTH) and β-lipotropin (β-LPH) has been elucidated by cloning and sequencing a cDNA copy of its mRNA (1), which was purified from the bovine pituitary intermediate lobe (2). Subsequently, the entire bovine (3), human (4-6) and mouse genes (7) encoding the ACTH-β-LPH precursor (designated alternatively as preproopiomelanocortin) have been isolated. They exhibit essentially the same structural organizations, consisting of three 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.

The production of ACTH and related peptides in the pituitary is regulated negatively by glucocorticoids and positively by corticotropin-releasing factor (8). Studies with the rat and a mouse pituitary tumor cell line in culture have shown that the cellular content of ACTH-β-LPH precursor mRNA is depressed by glucocorticoids, indicating that at least

© IRL Press Limited, Oxford, England.

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

part of the hormonal effect is exerted at the pre-translational level (9-11). Deletion mapping studies by mammalian-cell transfection and cell-free transcription approaches (12,13) have indicated that the TATA box region (14) is required for the accurate and efficient transcription of the human ACTH-β-LPH precursor gene and that removal of the sequence lying between 53 and 59 base pairs (bp) upstream of the capping site increases the transcriptional efficiency.

In the present investigation, we have determined the complete nucleotide sequence of the human ACTH-β-LPH precursor gene, which consists of 7665 bp including 1071-bp exonic sequences. Characteristic sequence features of this gene and its flanking regions are discussed.

MATERIALS AND METHODS

The plasmid pHAL1 containing the entire human ACTH-β-LPH precursor gene (4,15) was used. Plasmid DNA was isolated after induction with chloramphenicol as described by Kupersztoch and Helinski (16). Restriction endonucleases were purchased from Takara Shuzo Co. (Kyoto, Japan), Bethesda Research Laboratories (Rockville, USA) and New England Biolabs (Beverly, USA); reactions were conducted under the conditions recommended by the suppliers. Separation of restriction fragments was performed by electrophoresis on poly-

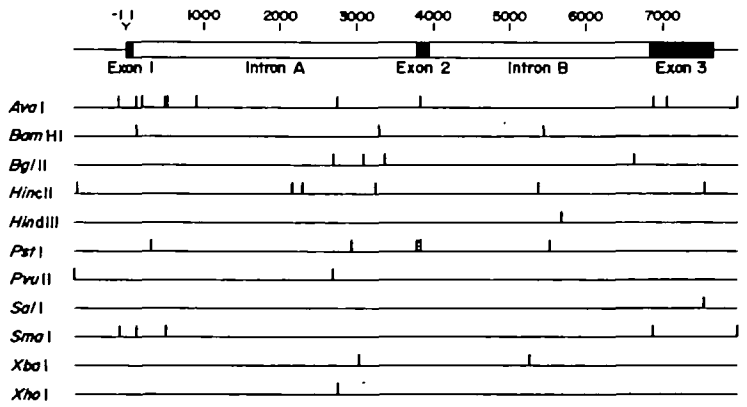

Figure 1. Restriction endonuclease sites in the human ACTH-β-LPH precursor gene region. The sequenced region, together with nucleotide numbers (see Fig. 2), is shown above the restriction endonuclease map; the exons are indicated by closed boxes, the introns by open boxes, and the flanking regions by lines. The direction of transcription is from left to right. All existing sites for the six-base-specific restriction endonucleases shown are indicated by vertical lines.

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

```
                                                  CTGCTCTTCACAGCATCACCCTCTCCCCATTTAATGGTTTAGGTTAACAGGACTTTTTCCTTGAGGCTTGGGACACGGAA -601
GGGAGCCTCCCCTAAACCAGGCCCTTGGAGAGCAGGCCCCAGGGGAGCAGTGCAACTCACCTTCACACCCACAAGACGGCTCCTGACTTCTGCTCCCTCC -501
TCCCCTCCCCAAAGTGGAACAGAGAGAATATGATTCCCCACGACTTCCACATCACAGTTTCCAAACAATGGGGAAATCGGAGGCCTCCCCGTGTGCAGAC -401
GGTGATATTTACCGCCAAATGCGAACCAGGCAGATGCCAGCCCCAGCACGCACGCAGGTAACTTCACCCTCGCCTCAACGACCTCAGAGGCTGCCCGGCC -301
TGCCCCACACGGGGGTGCTAAGCCTCCCGCCCGTTCTAAGCGGAGACCCAACGCCATCCATAATTAAGTTCTTCCTGAGGGCGAGCGGCCAGGTGCGCCT -201
TCGGCAGGACAGTGCTAATTCCAGCCCCTTTCCAGCGCGTCTCCCCGCGCTCGTCCCCCGTCTGGAAGCCCCCCTCCCACGCCCCGCGGCCCCCCTTCCC -101
CTGGCCCGGGGAGCTGCTCCTTGTGCTGCCGGGAAGGTCAAAGTCCCGCGCCCACCAGGAGAGCTCGGCAAGTATATAAGGACAGAGGAGCGCGGGACCA   -1
AGCGGCGGCGAAGGAGGGGAAGAAGAGCCGCGACCGAGAGAGGCCGCCGAGCGTCCCCGCCCTCAGAGAGCAGCCTCCCGAGACAGGTAAGGGCGCAGCG  100
TGGGGGACCCGTGCTCTTTCCCCGGGATCCCCTGTCCCCGTCCTCGCGATGCAGTCGGCCGGCTCCGGCTCCGAAGGCGGACCTGGGCGCCTCTGGCTCT  200
CCGCGGTCCCGAGTTCTCGACAAACTTTCTGCGCCGACTGCGGCATGAGAAGCCGCCAGTAGCTGAGCTGGAGGGCCCACGTCCGGCCCCTGGGCGGACG  300
GCCGCGAAGCTGCAGGCGCTGTCTCCAGGGAGCCGGCGGCCTCCTCTCCCCCAGGGGCTCGCGGCGGTCCGGAGGCTCCGAGAGCTTGCTAGGAGGTCTT  400
GGGACAACCCGGTCTTTTTTTTTTTTTTTGAGACGGAGTTTCGCTCTTGTTGCCCATGCTGGAGAGCAAAGGGGTGATCTCTGCTCACCGCAACCTTCGC  500
CTCCCGGGTTCAAGCGATTCTCCTGCTTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCGCCACCACGCCCGGCTAATTTTTGTATTTTTAGTAGTGA  600
CGGAGTTTCTCCATGTTGGTCAGGCTGGTCTCAAACTCCCGACAACAGGTGATCCGCCCGCCTTGGCCCCCCAAAGTTCTGGCATTACAGGCGCGAGCCA  700
CCGCCCCCGGCCAGCCCGGTCTTTTAGTATCTCTTGCTCCCAGTTTCCAGGATAGGTGTCACATCTTGAAAGTCAAATTCCATACACGCTATCGCAAATT  800
AATGTTGGAAACGGGGCAGCAGAGAAAAGGATAAAAGTCATAATGAACGCCCTGCCTTCCGGATTTTTTCGGATTCAGACCCCTGAATCCTTGTTTCCTT  900
GCCCACCTTAGCGCACCCGAGGTGGCCGCGCTATGATAATTACATGATAACTGGGTCAATTACAATGCAGAATAGTTGGGTCTCTTCTCTCCAAGACCTA 1000
GCTGGGGTTAAAAACAGGTGGCCGGGGCGGGAGCTGTCCTAGATCCTGAAACGCACTGTCTAGTTTCGGATGCCCTCAACAGAACCGGGGTGGACGGTTT 1100
ATGGCGCAGATCCTGGGTTGAGGGCACGGGCAGCCATTTGGAATGATCAAGGCTCAGGTAAGGGGCGTTTCCAGCGAAGGAGAGACAGTCCACTTGGCAT 1200
TTGGATTCCCCAAATTCTTCATGTTTAAATGGGGCAGGGAGGGTTCTTACAGAATGGCTGGAAGGAGCCAAGGAAAATAAAAGTGTGTGTGGATTTTTTT 1300
TGTGTGTGTGTCAGTTTATAAACTCTGCACAGATTATGGCCACTTTAATGACTTACTGTTCCTTTGATGCTTTTGTTATAGGACTCGATGCATGTATGTC 1400
ATGGTGTAAGGACAAAACTCGGCCCCTGTGCTCCTCTAATCTTTACAAAAGGTCATGGCCAGCGTGCAGTTTTACAGTAACAAGCAAAATGATTTGTTGA 1500
GCTCATAGAGAGCCCCTCACACCTATGAAGTTCTAATAAGTGTAGTTCTACTATAAAGTTAATCTCAGGATGAGCAAATTTCAAGTTTCTATTTTTCCAG 1600
AGCTTTCCATTTTTGGATTATAATACTTTCCCTACTTAAAAAAGCACAACATTTGATATTTCCCCAATAATTTGTTGCTTTAAAAATGACACAAAAGGTA 1700
CTATTTGTTCATTGTAGAGAACTGAAAATACACATAAGCAAATACACATACACATAAGCAAAATATACAATACAAACACAAGACCATCTTTCAGGGAAGA 1800
ATCTGAAGTTTTAGCAATAGCAGCCATCTAACCAGTTTAGCAACAGAATATAAGCTCTGAGAGGGTGGGAGTGAATATGTTACCACATTGTACAACACAG 1900
CACATAGGGCATAAGGAGGGGAAATGCTCTCTGGGGCTTTCCAGGAAGGCCTGAAGTCATTGCTTCTAGCAAATGGAAATCACTCCAGAGTAGTTATCTT 2000
TGACAAGAATTGAAATATAATTGAGGGAACTATCAGACCTGTAAGATTTTGTTTTTTCCTTTACTAATATGTTACTTTACATTTGCATTTGGTGACATAC 2100
GTAACTACCATTTTTCTGTGACTGTAACATCTGGGCATTTTTCAGAGCTAAATGTGCTATGGTCAACTTGGAGCTTTAATCTAATTGCCTGGTCCACCAA 2200
GTTCTGGCTGTGTACTTGAATAGATCACTGGCAGGGTACAATGGGAACAGCCTGTCCCTTGGAGCCAGGAGAGGACACCAAGGTTGACCAAAGCTCGTTC 2300
AGTTGCCCCTTTAGCCGAAGCGCACCTGGGCCAGTCACTGGCTGCCAGTGCCATCTAATGGCTGCTCTGAAAATGCTCAGCCTTGCCCGGCAACCCTTCA 2400
GAAGCTAGCACCGTGCAGGCCCAGCGCCTGGGGAATAGGGCGAGGGTGGGGTAGAGAGAAGGAAGTGGCCTCCTGAAGTAGAAATCAGCGCTTCAGAGGA 2500
CTTTCACTTCCAAAGCCTCCCCTATATAAAAAAGATTTGGCCCACGCCTCCCCAAATGAGAGATTTATTTTAGGCAAACTTATTTTAAAATGCCAGCGTT 2600
CATTAGGAGTGACAAGACACTTAGTCATCCACGCTTTAATGTGAATTACTTTTCTCATCTAATTACATTTCTTTCTAGCAGCTGGCTGAGAAGATCTTCT 2700
GAAATCCAAAATGATTGTAGGGTTGGCGGTGAGCTGATCTCCGGCCTCGAGGTGGCTTCAGGGGGCCCACCTGGTTAAGGGAAATTTGGCAGTGCGAGGG 2800
TAGTGCTGGAGAGAGGGGTGGGTACAGGGGGCTAGGGGCACCATGGATGCCCCCTCCTTACTGTCCCCTGGTGTCTTGACCTCAGCTTCTGCCCACAGGC 2900
ACTTGCTGGATTCTCCAAAAGTATCTGCAGTGGCTGTTCCACCAGGAGGTAATTCCCTTCTGGTCTCTTTCCCCTCCACATCTGCATCCTCTTCAAATCC 3000
```

Figure 2

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

TGCCATTTCAGACCACATTTGAGAGCTCTAGAGAACAAGACATCTGACACGTGACGTGTCCAGAAGATGAGCCAGATTTCAAAGAACTGAGATCTGCTTT 3100
AAAAACGAAGCTCTCCAAAGTTACTGGAGTCTGGGTAATAGTGATCACCAGAGTAATTTGTGTGCAGGACATCAAATCAGGCTGCTCGAAATGCTGCCTA 3200
AATTGGCCAGTGGTTTTATTTGCTTTTCTGTCAACCTAATATTCATAGGAAATAGAGTTTCAGAGGAATGATAGGATCCTGGTGGAATAAAAAGGGAAAA 3300
GACCATCTTGAGCAGGAGTTTCAGGGTCCTCCGTTTTTCCCAAGTTACTTTCACTCCTGAGATCTTGCATGTTAGAACTACAGCTTAATGTAGTGAAATA 3400
GGAAAGTTCTCTGTTAGGAGCTTAGCCTTACCTTGTCATGGACATTAAAGTAATTGTCTCTCTTTGGGCTTCAATTTTCCCATCTCTCATGGGAAGGGCT 3500
GAACCAAGCAATCCCCAAAATAGCTTCCAGCCTTAACCTTTTTAGGGGTCTCGTTTAAATAGAAGATAACAGGGAAATGGTCACAGTTTACCCAGGTCCA 3600
TTCCCTCCTCCTTATCACAACTTATACCACCGCTGTACTGCACACCTCCTTTCTCAGCATTGCTGCTGTCCTTAAAATGCCTTTAACTCCACAAGAGAGT 3700
GTGTTGTTAATGTTGGCTCAAGGTCCTTCCTGGTGAGTGGCCAACATTGTTTTGCTCCTTGCAGGGGTCCCACCAATCTTGTTTGCTTCTGCAGAGCCTC 3800
*** *** ***
AGCCTGCCTGGAAGATGCCGAGATCGTGCTGCAGCCGCTCGGGGGCCCTGTTGCTGGCCTTGCTGCTTCAGGCCTCCATGGAAGTGCGTGGCTGGTGCCT 3900
GGAGAGCAGCCAGTGTCAGGACCTCACCACGGAAAGCAACCTGCTGGTACGTGGGCCATGACTGCCATCTTGGCTTAGACATTAGATGGGACTGGAGCTG 4000
GGAAAGCTCAAAAGAAAAGGGTGTGGGGAAAGGGAAATTCATTCCCAGTGATAGGCGTGATTCAATCCAGGGCAGGAGCAAAACTTTGCAGTGAAGTAAG 4100
AAATGGGAGAAGAAATCAGGGAAGGAAGCAGCTTCAGGGAGAGGGGTTGAGTCCACAATTTCTGCTTGGTTATCCTTACTTCTTGCCCCATCTTTTATGG 4200
AGACCTTGAACCCTTTAAGCTAGAGATGGTGCTATAAGAGCAATAATGGACCCCTCAATCTATTCTGTACTTTACATCTTTAGCTTCCCAAACTATTCCT 4300
TTTTAAGAAGCTCATATCACTTGCCATTTTCATTCCATATTTCTTACCCTTTTATCTACTACCGGTTGCAAAACCAGCCAGGTAGTTCTTCAAATCATCT 4400
CTGGAAGAAGGAAAAACCAGGGGCCCTTTTTTTTTTTTCTTTAATTGGTGCCAAATGTCTCATGTTTATTCTGGAGGACTGGCCTTCTGCTGTGTTCCTC 4500
TACAGTCTTTCCAGAGCATGTGAAGGCCTTTGCATCAGGCAGGAGCTCCCTCCAGGTCACCACAGGGTGTATGTATCTGCCTGTGGGGGGTGTGTGTGTG 4600
TGTGTGTTGGGGGGCATAAATGAGTAATGATGCCAAATCCAGAGATTAAAAGGCACACTGAGACCAGGCGAGATGGCTCATGGCTGTAATCCCAGCACTT 4700
TTAGATGCTAAGGTGGGAGGATTGCTTGAGCCCAGGGATTCAAGACAAGCCTGGGCAACATAGTGAGACCTCCACTTCTACAAAAAATAAAAAAGTTAGC 4800
CAGATGTGGTGGCATGTGCCTGTAGTCCTAGCTACTTGGGAGGTTCACTTGAGGCCAGGAGTCTGACGACACAGTAAGCTATGATCACACCATTGCACTC 4900
CAGTCTGGGTAACAGAATGAGACCTTGTCTCAAAACAAAACAAAATGAAACAAACAAACAAACAAACCCCCATACTGTTAGTGTCAGTGACCGGAATTTT 5000
AATCTTGTTGCCATCACCTGGCAGGTGCTGAGGGTGGAATGTACATAACTACATTCTGTGTATTTTGTCAATGCAGAAGCTGAGTTAAGGTGAAGATAGA 5100
ATGAGGTCCTCAAAGACACAGACCAGTTTTCATGTGTAATATAAAATAGAAACAAAGAGCCCAGGGGATTCTGTGAGTTCCAGTTTGGAAAGACCCAAGA 5200
GTCTCTTGACTTGAGACACCCACAGCACAGCTCACCAGGGAGGGTGCACTGGACACAGTCAGGACCCATGGGTTCTAGACCCAGTTTTGAGGTGTGGGAC 5300
CTTGACCAGGTCCTATCACCTCTCTGAGTCTCCTGTTTCACTATCTGTCCACGGGAGGGGAGTGTAAATTAGTTTTTTCCATTGTTAACGTTCCACAGAG 5400
TTGTAATTCTGAACACCTGGAGTAGGCAATGTCCAGCTCAACAGAGTGGGTAGGATCCTTTTATTTTCTCCTTTGCTATTCCCAAGAAAGAGAGCAGCCA 5500
GTGAGCTTTTCATCTTTTTATCACTGAAAACTCAAGGCTGCAGCCTATGCAGCCATTTTCCTAAGCTAATATGTACCACAATAGAGTCCTCTAGGGACAA 5600
GGAGCAGAGACACAGGTTCCACAGACGGTGCAATGGAAATAACGCTAGCTTTCCACCCCTCCCTCCAGTCAGAATGAGATTACAGGGAAATAAGCTTGCC 5700
CCAGAGCTCACTGGGGGATCTCTCAGAAATCAGCTCAGAAGTCGTGAAAGAACCAAGGTGCAGTTTTGGAGGCTTAGTGCAGAGATGGAGCTGGGGTAGG 5800
GCATAAAGTAGGTTTTCCATCACTGAGGTAAGGTTGAGGCATTATTTTTTATTTTTTGTTTATTTATTTATTTTTTTGAGACGGAGTCTCGCTCTATCAC 5900
CCAGGCTGGAGTGCAGTGGCGCGATCTCCCCTCACTGCAAGCTCCACCTCCCAGGTTCACACAGGTTGAAGCATTATTAAAAATATGTTTAAAAATATGG 6000
GCCCTAGTAGCCAGACTTCTATCACCTGGAGAGATTATCCCCCAAATTTCAGCCCCACTCCCCTCCTGGACTTGAATTAAACCATATGTATTTATTCAAT 6100
ATTCTTTTTATTTATTTATTTATTTTTTTGAGACGGAGTCTTGCTCTGTTGCCCTGGCTGGAGTGTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAAC 6200
CTCTACCTCCCAGGTTCAAGCGGTTCTCCTGCCTCAGGCTCCAGAGTAGCTGGGATTACAGGCGCCCGCCACCACACCCAGCTTATTTATTTATTTATAC 6300
TAGAGATGGTATTTCACCATAGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCATGTGATCTGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTATAGGTG 6400
TGAGCCACCATGCCCGGCCCTCAATATTCATTAAGTGCCAACAACTACCACCCGTCTGCCTTTCTTGGAGCCACTCCTTTATGTCAGGCATATGACAGTA 6500
AGACTTTGGTCCTGTTCACAAAAGCTAGGGGTGGCTAGATGGCTAGACAAACCATGGAATGGGATGGGAAGTGTGTTGCAGTTGCCAGCAGAAGCATGAA 6600
GGGGATGGGACAAAAGAGGCGGTGGCAAGATCTTAGATGCCCACGAGTGCCAAGAAAGCAGGTGGGCAGACCTGCTCTGTAGGGAGGCCTCGACGCTTGA 6700

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

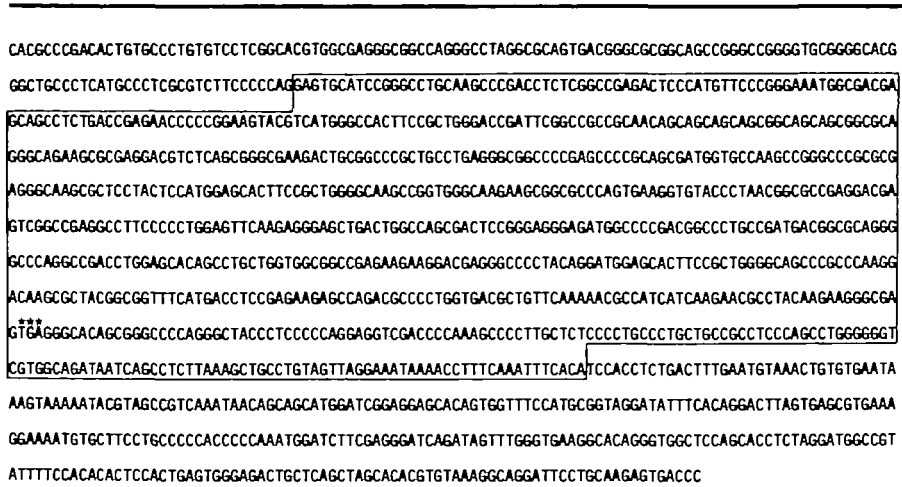

Figure 2. Nucleotide sequence of the human ACTH-β-LPH precursor gene and its flanking regions. The sequence of the message strand is shown. Nucleotide residues are numbered in the 5' to 3' direction beginning with the putative capping site, and the nucleotides on the 5' side of residue 1 are indicated by negative numbers; the number of the nucleotide residue at the right end of each line is given. The sequences of the exons are boxed. The Alu family sequences in the introns are underlined; for the orientation of these sequences, see ref. 15. The TATA box (14) and the CAAT box (18) are overlined. The translational initiation and termination codons are marked with asterisks. The sequences of the exons and adjacent regions, the 5'-flanking region and the Alu family members were reported previously (4-6,15).

acrylamide gel in 50 mM Tris-borate buffer (pH 8.3) containing 1 mM EDTA. Fractionated DNA on polyacrylamide gel was extracted by maceration. 5'-End labelling of restriction fragments and DNA sequencing were carried out by the method of Maxam and Gilbert (17).

RESULTS AND DISCUSSION

Figure 1 shows a map of the sites for various six-base-specific restriction endonucleases present in the 8658-bp human genomic DNA segment that was sequenced. All these sites were confirmed by digestion of the plasmid pHAL1 with the respective restriction endonucleases.

Figure 2 presents the nucleotide sequence of the entire human ACTH-β-LPH precursor gene and its flanking regions. Most of the nucleotide sequence shown was determined on both strands, and all sites used for end labelling were overlapped by at least one different sequence. On the basis of S1 nuclease mapping analysis and the fact that eukaryotic mRNAs generally start with an A residue (reviewed in ref. 19), the putative capping site has been

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

assigned to residue 1 (12,13); the previous assignment of the capping site to the A residue at position -1 (4-6) was based on comparison with the bovine gene sequence (3,20). The putative poly(A) addition site has been assigned to residue 7665 by comparison with the bovine gene sequence (1,3). On the basis of these assignments, the human ACTH-β-LPH precursor gene is 7665 bp in length and has the following structure (in the 5' to 3' direction): exon 1 (86 bp) - intron A (3708 bp) - exon 2 (152 bp) - intron B (2886 bp) - exon 3 (833 bp). The 5'-flanking sequence of 680 bp and the 3'-flanking sequence of 313 bp are also presented.

Previous work by our group (15) has shown that intron A contains one member of the Alu family of dispersed repetitive sequences (21,22), while intron B contains two Alu family sequences and a sequence corresponding to the 3' one-third of an Alu family member. The locations of these repetitive DNA sequences are indicated in Fig. 2. Computer analysis detects no Alu family sequences other than those mentioned above, nor other repetitive DNA sequences such as the EcoRI 340-bp dimer repeats (23), the HindIII 1.9-kilobase-pair family (24) and the Hinf family (25). Furthermore, no significant sequence homology is noted between intron A and intron B, except for the Alu family sequences.

Table 1 shows the base composition and the dinucleotide frequency for the different regions of the human ACTH-β-LPH precursor gene and its flanking regions. The A + T content of the intronic sequences (52.0%) is higher than that of the exonic sequences (33.1%). The G + C content of the 5'-flanking region (61.5%) exceeds that of the 3'-flanking region (49.8%). The frequency of the dinucleotide sequence CpG is higher in the protein-coding, 5'-untranslated and 5'-flanking regions than in the other regions. A survey of other mammalian genes has shown that this rare dinucleotide sequence is asymmetrically distributed, occurring more abundantly in the protein-coding and 5'-flanking regions (26). Intron A contains two sequences of alternating purine and pyrimidine nucleotides, that is, $(TG)_4$ and $(TG)_5$ (residues 1284-1291 and 1301-1310) at a distance of 9 bp, and intron B contains one such sequence, that is, $(GT)_9$ (residues 4590-4607).

Three open reading frames that could encode a reasonably large translation product (99 amino acids or more) are found. They start with the ATG triplet at residues 456-458 in intron A and at residues 4246-4248 and 6637-6639 in intron B and could encode translation products of 99, 125 and 138 amino acids, respectively. The translation initiated at the triplet composed of residues 6637-6639 in intron B would proceed into exon 3 using

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

Table 1. Parameters of base distribution in the human ACTH-β-LPH precursor gene and its flanking regions. The base composition (in percent) and the frequency of dinucleotides have been calculated separately for the different regions indicated. The values of dinucleotide frequency are shown by the ratio of the observed number to the number expected from the base composition.

		Protein-coding region	5'-Untranslated region	3'-Untranslated region	Intron A	Intron B	5'-Flanking region	3'-Flanking region
Base composition (%)	T	12.5	5.7	19.5	27.1	26.5	17.6	22.4
	C	32.7	33.0	35.4	24.0	22.9	36.5	23.3
	A	19.7	23.6	20.7	24.8	25.6	20.9	27.8
	G	35.1	37.7	24.4	24.1	25.0	25.0	26.5
Dinucleotide composition	TpT	1.04	0	1.13	1.24	1.21	1.51	0.96
	TpC	0.92	2.06	0.89	0.99	0.94	1.01	0.98
	TpA	0.36	0	0.91	0.71	0.70	0.64	0.72
	TpG	1.43	0.90	1.16	1.05	1.15	0.93	1.35
	CpT	1.47	2.57	1.33	1.13	1.13	1.10	1.04
	CpC	1.06	1.32	1.37	1.22	1.30	1.18	1.24
	CpA	0.99	0.49	0.92	1.16	1.33	1.06	1.29
	CpG	0.79	0.85	0.28	0.47	0.25	0.61	0.41
	ApT	0.81	0	0.61	0.84	0.83	0.76	0.88
	ApC	0.72	0.25	0.42	0.75	0.83	0.73	0.94
	ApA	1.06	0.69	1.71	1.22	1.04	1.11	1.00
	ApG	1.30	2.05	1.46	1.21	1.30	1.44	1.17
	GpT	0.66	0.45	0.64	0.77	0.84	0.70	1.08
	GpC	1.13	1.08	1.07	1.06	0.96	0.94	0.88
	GpA	1.19	1.62	0.61	0.95	0.98	1.07	1.00
	GpG	0.87	0.54	1.55	1.25	1.22	1.25	1.05

an illegitimate reading frame and would terminate at the triplet composed of residues 7051-7053.

Figure 3 shows the alignment of the nucleotide sequences upstream of the capping site of the human, bovine and mouse ACTH-β-LPH precursor genes. The percent nucleotide difference along the 5'-flanking region between each pair of the aligned sequences is illustrated in Fig. 4. This analysis, in conjunction with inspection of the aligned sequences, reveals that the region corresponding to positions -353 to -267 (for the numbering of positions in the aligned sequences, see the legend to Fig. 3) is more highly conserved than the other regions and exhibits almost no deletion/addition. This highly conserved region represents the sequences lying 304-219 bp, 344-258 bp and 298-213 bp upstream of the capping site of the human, bovine and mouse genes, respectively. Computer analysis shows that this region contains two overlapping sequences of 14 bp (positions -290 to -277) and 14-15 bp (positions -284 to -270) that are homologous with the sequences TTAA$^{A}_{G}$TAAGTTT$^{T}_{A}$T and AAGTGGTTTCCTGAC located 198-185 bp and 149-135 bp upstream of the mRNA start site of the mouse mammary tumor virus long-terminal repeat (27,28), respectively. Such overlapping homologous sequences (TTAATCAAATCCTTCCTTTC) are

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

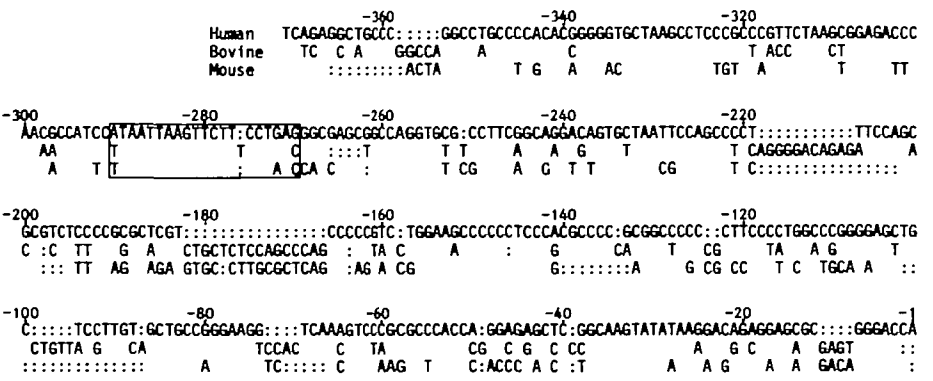

Figure 3. Alignment of the 5'-flanking sequences of the human, bovine and mouse ACTH-β-LPH precursor genes. The sequences of the message strand are shown. The absence of a nucleotide in the bovine or mouse sequence means that it is identical with the human sequence. The presence of a colon in either sequence indicates a gap; gaps have been inserted to achieve maximum homology. The positions in the aligned sequences including gaps are indicated by negative numbers beginning with the position immediately preceding the capping site. The segment comprising homologous sequences shared by the 5'-flanking regions of other glucocorticoid-regulated genes is boxed. The bovine and mouse sequences have been taken from refs. 3 and 7, respectively, except that the most upstream 159-bp sequence of the bovine 5'-flanking region was determined in the present work.

also found 198-179 bp upstream of the capping site of the rat prolactin gene (29). Furthermore, the sequence corresponding to positions -284 to -270 exhibits homology with the sequences GAGATCTTGCGTGAC and GTGGTATTTTCTGGC (or TATTTTCTGGCTGAC) lying 238-224 bp and 235-221 bp (or 231-217 bp) upstream of the capping site of the rat (30) and human growth hormone genes (31), respectively, and the sequence corresponding to positions -290 to -277 is homologous with the sequence TTAAATGATTTCTA located 101-88 bp upstream of the capping site of the chicken ovalbumin gene (18). In addition, there is a 10-bp sequence (positions -333 to -324) that is fully conserved among the 5'-flanking regions of the three ACTH-β-LPH precursor genes and that is homologous with the sequence AGTCAGCCTC lying 237-228 bp upstream of the capping site of the rat prolactin gene.

The expression of the mouse mammary tumor virus gene (32,33) as well as the growth hormone gene (34,35) is positively regulated by glucocorticoids, while the production of prolactin (36), like that of the ACTH-β-LPH precursor (9-11), is negatively controlled by the steroids. The expression of the chicken ovalbumin gene is also stimulated by glucocorticoids administered

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

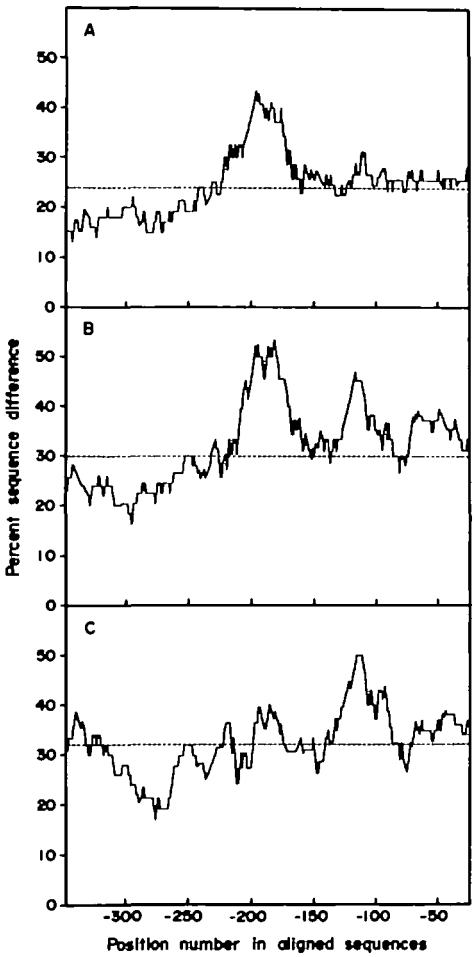

Figure 4. Percent sequence difference along the 5'-flanking region of the human, bovine and mouse ACTH-β-LPH precursor genes. Comparison is made between the human and bovine sequences (A), between the human and mouse sequences (B) and between the bovine and mouse sequences (C). The percent sequence difference of a segment comprising positions $\underline{i}$ - 24 to $\underline{i}$ + 25 has been plotted against $\underline{i}$, where $\underline{i}$ represents the position number in the aligned sequences (see Fig. 3); gaps have been counted as one substitution regardless of their length. The dashed lines indicate overall percent sequence differences throughout the whole length of the 5'-flanking sequences compared.

after estrogen treatment (37). It has recently been shown that the "glucocorticoid response element" in the mouse mammary tumor virus long-terminal repeat is located within a segment extending between 449 bp and 109 bp upstream of the mRNA start site and that this segment contains specific binding sites for purified glucocorticoid receptor protein (38,39). There is also evidence indicating that the 202-bp segment preceding the mRNA start site of this long-terminal repeat is competent to confer glucocorticoid responsiveness (40). Thus, the region responsible for the hormone sensitivity of mouse mammary tumor virus gene expression includes the above-mentioned sequences homologous with the conserved upstream sequences shared by the ACTH-β-LPH precursor gene, prolactin gene, growth hormone gene and ovalbumin gene. It

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

is tempting to speculate that these conserved sequences may be involved in the modulation of expression of the glucocorticoid-regulated genes.

Virtually all splice junction sequences found in nuclear and viral protein-coding genes conform to well-defined consensus sequences, that is, $\begin{matrix}C\\A\end{matrix}AG/GT\begin{matrix}A\\G\end{matrix}$AGT for donors and $(\begin{matrix}T\\C\end{matrix})_{11}N\begin{matrix}C\\T\end{matrix}$AG/G, where N can be any base, for acceptors (41). In the processing of the RNA transcript of the mouse β-globin gene (42) as well as the chicken α2-collagen gene (43), excision of intronic sequences occurs in a stepwise fashion; at each step only a portion of an intronic sequence is removed. It has also been reported that alternative splicing pathways generate protein polymorphism, for example, in mouse immunoglobulin heavy chains (44,45), chicken ovomucoid (46) and human growth hormone (31). The DNA sequence of the human ACTH-β-LPH precursor gene was therefore searched for potential splice sites other than the legitimate ones. Donor sites are defined as sequences with two or fewer mismatches from $\begin{matrix}C\\A\end{matrix}AG/GT\begin{matrix}A\\G\end{matrix}$AGT, not including mismatches with the underlined GT. Acceptor sites are defined as sequences with two or fewer mismatches from $(\begin{matrix}T\\C\end{matrix})_{11}N\begin{matrix}C\\T\end{matrix}$AG/N, not including mismatches with the underlined $\begin{matrix}C\\T\end{matrix}$AG and with no AG dinucleotide less than 13 nucleotides upstream of the conserved $\begin{matrix}C\\T\end{matrix}$AG. In addition to the legitimate splice sites, 12 potential donor sites and 13 potential acceptor sites were found by this search; the donor sites are located at residues 646-654, 1280-1288, 1391-1399, 2638-2646, 2946-2954, 3447-3455, 3730-3738, 4379-4387, 5087-5095, 5805-5813, 5825-5833 and 6659-6667, and the acceptor sites at residues 340-355, 517-532, 729-744, 1586-1601, 2664-2679, 3531-3546, 3643-3658, 3750-3765, 4491-4506, 5654-5669, 6199-6214, 6223-6238 and 7524-7539. All these potential splice sites lie within the introns, except the acceptor site at residues 7524-7539. Previous work by our group has shown that a human ectopic ACTH-producing tumor contains two mRNA species hybridizable with a bovine ACTH-β-LPH precursor cDNA probe and that one of them, representing a minor species, exhibits a size larger than that of the pituitary ACTH-β-LPH precursor mRNA by about 200 nucleotides (47). It is possible that this larger mRNA species may have been generated by splicing at alternative sites.

ACKNOWLEDGEMENTS

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.

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

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. Notake, M., Tobimatsu, T., Watanabe, Y., Takahashi, H., Mishina, M. and Numa, S. (1983) FEBS Lett. 156, 67-71.
8. Yates, F.E. and Maran, J.W. (1974) in Handbook of Physiology, Greep, R.O. and Astwood, E.B. Eds., Vol IV, pp. 367-404, American Physiological Society, Washington, D.C.
9. Nakanishi, S., Kita, T., Taii, S., Imura, H. and Numa, S. (1977) Proc. Natl. Acad. Sci. USA 74, 3283-3286.
10. Nakamura, M., Nakanishi, S., Sueoka, S., Imura, H. and Numa, S. (1978) Eur. J. Biochem. 86, 61-66.
11. Schachter, B.S., Johnson, L.K., Baxter, J.D. and Roberts, J.L. (1982) Endocrinology 110, 1442-1444.
12. Mishina, M., Kurosaki, T., Yamamoto, T., Notake, M., Masu, M. and Numa, S. (1982) EMBO J. 1, 1533-1538.
13. Notake, M., Kurosaki, T., Yamamoto, T., Handa, H., Mishina, M. and Numa, S. (1983) Eur. J. Biochem., 133, 599-605.
14. Goldberg, M. (1979) Ph.D. Thesis, Stanford University.
15. Tsukada, T., Watanabe, Y., Nakai, Y., Imura, H., Nakanishi, S. and Numa, S. (1982) Nucleic Acids Res. 10, 1471-1479.
16. Kupersztoch, Y.M. and Helinski, D.R. (1973) Biochem. Biophys. Res. Commun. 54, 1451-1459.
17. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
18. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
19. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
20. Inoue, A., Nakamura, M., Nakanishi, S., Hidaka, S., Miura, K. and Numa, S. (1981) Eur. J. Biochem. 113, 531-539.
21. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
22. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature 284, 372-374.
23. Wu, J.C. and Manuelidis, L. (1980) J. Mol. Biol. 142, 363-386.
24. Manuelidis, L. (1982) Nucleic Acids Res. 10, 3211-3219.
25. Shimizu, Y., Yoshida, K., Ren, C.-S., Fujinaga, K., Rajagopalan, S. and Chinnadurai, G. (1983) Nature 302, 587-590.
26. McClelland, M. and Ivarie, R. (1982) Nucleic Acids Res. 10, 7865-7877.
27. Donehower, L.A., Huang, A.L. and Hager, G.L. (1981) J. Virol. 37, 226-238.
28. Fasel, N., Pearson, K., Buetti, E. and Diggelmann, H. (1982) EMBO J. 1, 3-7.
29. Cooke, N.E. and Baxter, J.D. (1982) Nature 297, 603-606.
30. Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4867-4871.
31. DeNoto, F.M., Moore, D.D. and Goodmann, H.M. (1981) Nucleic Acids Res. 9, 3719-3730.

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024

32. Parks, W.P., Ransom, J.C., Young, H.A. and Scolnick, E.M. (1975) J. Biol. Chem. 250, 3330-3336.
33. Ringold, G.M., Yamamoto, K.R., Tomkins, G.M., Bishop, J.M. and Varmus, H.E. (1975) Cell 6, 299-305.
34. Martial, J.A., Baxter, J.D., Goodman, H.M. and Seeburg, P.H. (1977) Proc. Natl. Acad. Sci. USA 74, 1816-1820.
35. Tushinski, R.J., Sussman, P.M., Yu, L.-Y. and Bancroft, F.C. (1977) Proc. Natl. Acad. Sci. USA 74, 2357-2361.
36. Tashjian, A.H.Jr., Bancroft, F.C. and Levine, L. (1970) J. Cell Biol. 47, 61-70.
37. Hager, L.J., McKnight, G.S. and Palmiter, R.D. (1980) J. Biol. Chem. 255, 7796-7800.
38. Chandler, V.L., Maler, B.A. and Yamamoto, K.R. (1983) Cell 33, 489-499.
39. Yamamoto, K.R., Payvar, F., Firestone, G.L., Maler, B.A., Wrange, Ö., Carlstedt-Duke, J., Gustafsson, J.-A. and Chandler, V.L. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 977-984.
40. Hynes, N., van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H. and Groner, B. (1983) Proc. Natl. Acad. Sci. USA 80, 3637-3641.
41. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
42. Kinnigurgh, A.J. and Ross, J. (1979) Cell 17, 915-921.
43. Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V.Jr., Pasten, I. and Benoist, C. (1980) Cell 21, 689-696.
44. Early, P., Rogers, J., Davis, M., Calame, K., Bound, M., Wall, R. and Hood, L. (1980) Cell 20, 313-319.
45. Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W. and Tonegawa, S. (1981) Cell 24, 353-365.
46. Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) Cell 21, 681-687.
47. Tsukada, T., Nakai, Y., Jingami, H., Imura, H., Taii, S., Nakanishi, S. and Numa, S. (1981) Biochem. Biophys. Res. Commun. 98, 533-540.

Downloaded from https://academic.oup.com/nar/article/11/19/6847/1067095 by guest on 23 April 2024