Structure of the rat pro-opiomelanocortin (POMC) gene

Jacques Drouin, Michel Chamberland, Jean Charron, Lucie Jeannotte and Mona Nemer

Laboratoire de Biologie Moléculaire des Eucaryotes, Institut de recherches cliniques de Montréal, 110 ouest, Avenue des Pins, Montréal, Québec H2W 1R7, Canada

Received 31 August 1985

The gene encoding pro-opiomelanocortin (POMC) presents unique regulatory features. In particular, glucocorticoids inhibit transcription of the POMC gene in the anterior pituitary, but not in the intermediate pituitary. In order to study the mechanism leading to transcriptional inhibition of POMC by glucocorticoid and the interaction of the glucocorticoid receptor complex with specific DNA sequences along the POMC gene, we have cloned the rat POMC gene and determined its structure. The gene is composed of three exons and appears to be present at a single copy per haploid genome. Besides the usual regulatory signals like 'TATA' and 'CCAAT' boxes, the upstream region contains sequences homologous to known enhancer sequences and to the glucocorticoid receptor binding site observed in glucocorticoid-responsive genes.

Pro-opiomelanocortin *DNA sequence* *Gene structure*

1. INTRODUCTION

Pro-opiomelanocortin (POMC) is the protein precursor to a variety of active hormonal peptides; they include adrenocorticotropin (ACTH), β-endorphin, β-lipotropin (β-LPH), and the melanotropins (α, β, γ-MSH) [1,2]. POMC mRNA is found in the anterior and intermediate lobes of the pituitary, in certain areas of the brain like the hypothalamus, the amygdala and the cortex, in the testes, ovaries, placenta, and in a variety of tumors [3–7]. POMC is differentially processed in the pituitary: ACTH and β-LPH are the major processing products in the anterior lobe whereas α-MSH and β-endorphin are more abundant in the intermediate pituitary [1]. The release of POMC-derived peptides from the pituitary is under multihormonal control [8]. In the anterior but not in the intermediate pituitary, glucocorticoids inhibit POMC gene transcription [9–11]; corticoliberin (CRF) stimulate POMC gene transcription [11]. Thus, the POMC gene is an interesting model system to study negative regulation of transcription by steroid hormones. We have reported the DNA sequence of the third exon of the rat POMC gene containing most of the coding sequences [12]. We now report the complete structure of the gene and DNA sequences around all 3 exons including 711 base pairs (bp) upstream of the site of initiation of transcription. The upstream DNA sequence will be useful to study by surrogate genetics the regulation of POMC gene transcription.

2. MATERIALS AND METHODS

Standard cloning techniques were used [13]. A rat genomic DNA library was constructed in bacteriophage λ1059 [14]. Genomic DNA was isolated from Sprague-Dawley rat livers and partially digested with *Sau*3A. The DNA was size-fractionated on agarose gel and the 18–22 kb fraction was ligated with *Bam*HI-digested λ1059 DNA. A library containing 1.4×10^6 recombinant phages was obtained and screened by hybridization with a nick-translated 1.6 kb *Xho*I-*Hin*dIII fragments containing the third exon of rPOMC [12]. Primer extension was performed as in [15]. DNA sequencing was carried out by the procedure of Sanger et al. [16], using M13 single-stranded DNA templates, and Maxam and Gilbert [17].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

3. RESULTS AND DISCUSSION

The complete POMC gene was isolated from a Sprague-Dawley rat genomic DNA library cloned in bacteriophage λ1059. A 1.6 kb *Xho*I-*Hin*dIII genomic DNA fragment containing the third exon of rPOMC [12] was used as probe. Four independent clones were isolated and characterized by restriction enzyme mapping. Two of these contain more than 6 kb upstream of previously reported rPOMC sequences [12]. All 4 clones have similar restriction maps when compared with each other and with 3 previously characterized rPOMC clones isolated from a Long-Evans rat genomic DNA library [12]. Fragments from the λ clones were subcloned in pBR327 to facilitate analysis. Exons 1 and 2 were localized by hybridization with a DNA probe (about 370 bp) made by primer extension on neurointermediate pituitary poly(A$^+$) RNA using a 118 bp *Hae*III fragment as primer (encoding amino acids 23–62 of POMC). Fig.1 illustrates the structure of the rPOMC gene: the

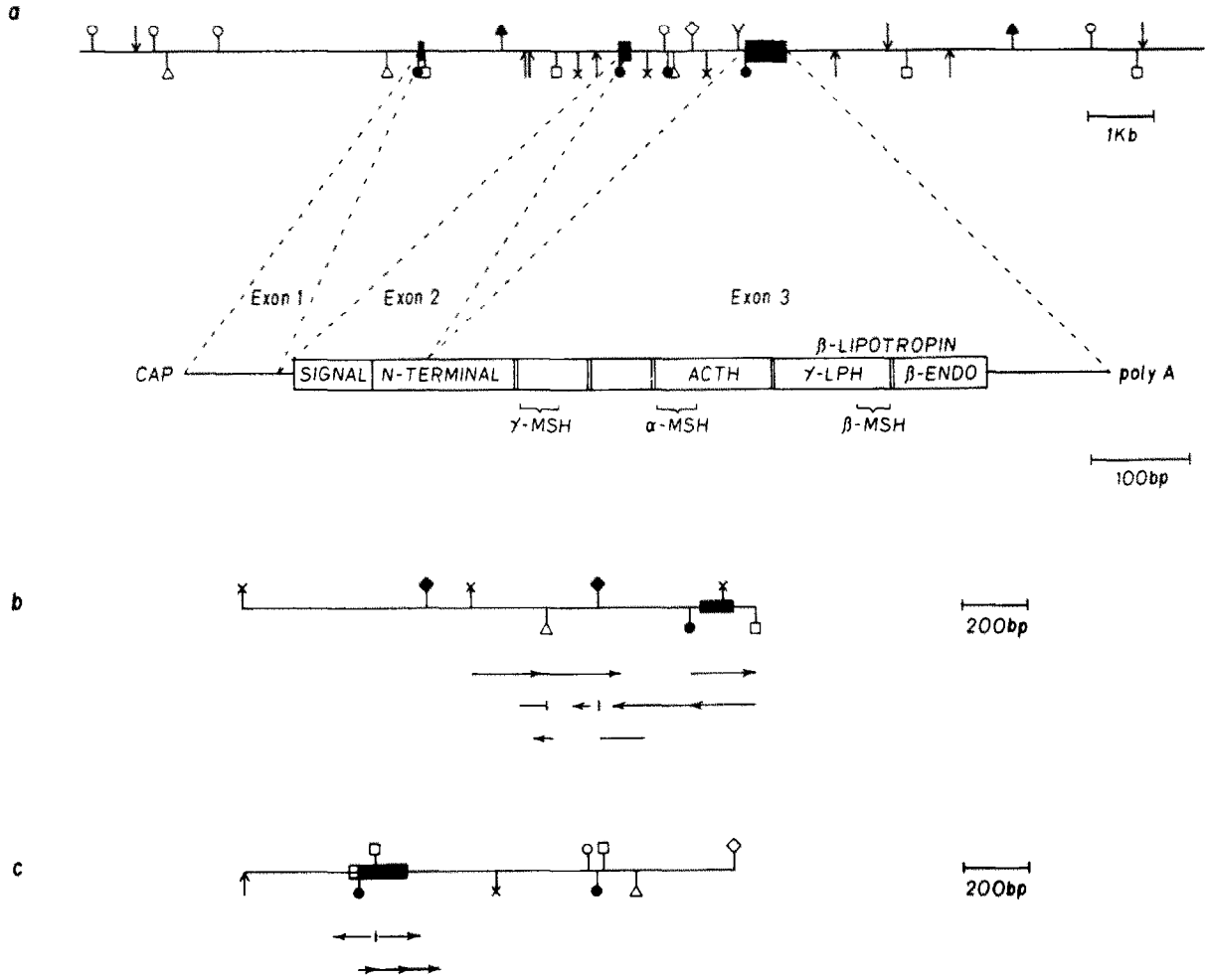

Fig.1. Structure of rat POMC gene. (a) Rat genomic DNA restriction map. Restriction enzyme cleavage sites: *Eco*RI (↓), *Hin*dIII (↑), *Xba*I (♀), *Bgl*II (△), *Pst*I (▼), *Bam*HI (□), *Acc*I (✕), *Hpa*I (♡), *Xho*I (Y), *Kpn*I (↑). POMC exon sequences are indicated by solid boxes on the restriction map; a schematic digram of rPOMC mRNA shows regions present in each exon. The translated portion of the mRNA is shown as an open box. (b,c) Strategy used for DNA sequence determination of exons 1 and 2, respectively. Arrows below the diagrams indicate the length and direction of sequence determination. Symbols as in a; and *Xmn*I (†), *Stu*I (♦), *Apa*I (♀).

restriction enzyme cleavage map is shown together with the position and size of exon sequences. The rPOMC gene spans about 6 kb and is divided in 3 exons like the human [18–21], bovine [22] and mouse [23,24] POMC genes. The first and second introns are about 3 and 1.8 kb in size, respectively. Repeated sequences are present in the first intron within the 1 kb *Hin*dIII fragment as both this 1 kb *Hin*dIII fragment and the overlapping 2 kb *Bam*HI fragment, but not the 1.6 kb *Bam*HI-*Hin*dIII fragment, hybridize to a smear in rat genomic DNA blots. Analysis by Southern blotting of rat genomic DNA isolated from various sources (livers of Sprague-Dawley, Long-Evans and NEDH rats, GH_3 (Wistar-Furth rat) and HTC (buffalo rat cells) suggests the cloned gene is unique (not shown). The discrepancy between these results and previous ones is due to contaminating plasmid DNA present in the genomic DNA preparation used in [12].

The DNA sequence of regions containing exons 1 and 2 was determined by the dideoxy chain termination method after subcloning appropriate fragments in bacteriophage M13 mp8 and mp9 [15]. The sequencing strategy is indicated by arrows in fig.1b and c. The sequenced DNA (fig.2) includes 711 bp in the 5′-flanking region which might contain important regulatory sequences. The exon-intron borders were assigned by comparison with a rat neurointermediate pituitary cDNA clone obtained in this laboratory; this cDNA clone includes the 2 exon-intron junctions and contains the last 21 bp of exon 1. Oates and Herbert [25] obtained a rat POMC mRNA sequence by specific priming on pituitary poly(A^+) RNA: both sequences agree except for the codon for amino acid $^{-}$21. A Tyr residue is also found at that position in mouse POMC [23,24], like in our rPOMC sequence. Oates and Herbert [25] also

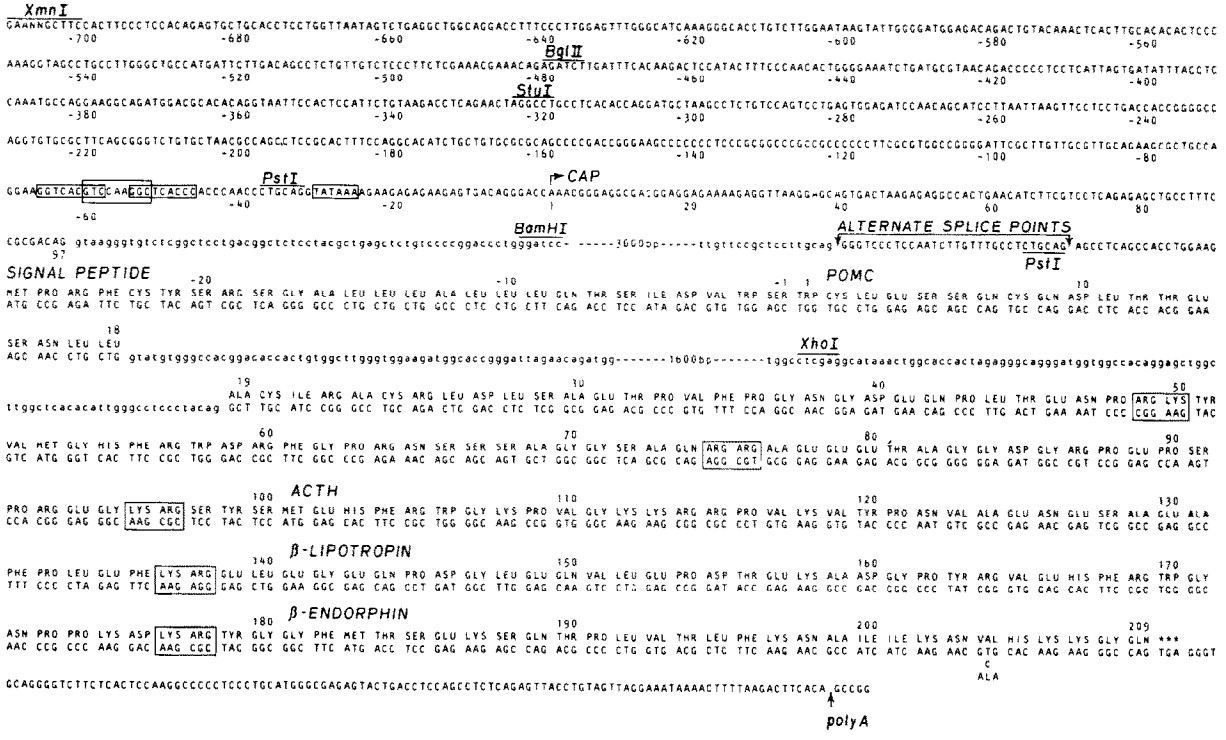

Fig.2. DNA sequence of the rat POMC gene. Intron sequences are shown in lower case. Amino acids are numbered positively for POMC and negatively for the signal peptide sequence. Similarly, nucleotides are numbered positively downstream from the site of transcription initiation (CAP) and negatively upstream. The TATA box and 3 putative CCAAT boxes are boxed. The alternate splice points used in pituitary [25] are shown at the 5′-end of exon 2. The exon 3 sequence is as in [12]; however, sequence analysis of a Sprague-Dawley rat cDNA indicates a difference at codon 204 of POMC between this strain and Long-Evans rat as reported [12].

characterized a POMC mRNA produced by alternate splicing: the additional 30 bp segment present in this mRNA is found at the 3′-end of intron A. The site of initiation of transcription (CAP in fig.2) was assigned on the basis of our primer-extension studies, of those of Oates and Herbert [25] and by comparison with the human, bovine and mouse POMC genes [18–24].

A perfect TATAAA sequence is found 30 bp upstream of the site of transcription initiation. Three sequences homologous to the GG$^{C}_{T}$CAATCT consensus are found between $^{-}$67 and $^{-}$47 bp. Such sequences were shown to be important in determining the site and efficiency of transcription initiation, respectively [26]. Enhancer sequences play an important role in transcriptional activation and, in some cases, confer tissue specificity [26–28]. Sequences homologous (7/8) to the SV40-type core consensus enhancer sequence [27], GTGG$^{AAA}_{TTT}$G, are found at $^{-}$357, $^{-}$246, $^{-}$188 and $^{+}$85 bp. In addition, sequences homologous (8/10) to another enhancer consensus characterized initially in adenovirus [28] are found at $^{-}$257 and $^{-}$66 and $^{+}$14 bp. Interestingly, SV40 and adenovirus-type consensus sequences are contiguous at positions $^{-}$257 and $^{-}$246 bp; they are both in opposite orientation to transcription. Two similar sequences are also found close to each other (9 bp away) upstream of the rat growth hormone gene at positions $^{-}$193 and $^{-}$176 [29,30]; in this case, they are both in the same orientation as transcription. The hexanucleotide sequence CCGCCC present in the SV40 21 bp repeats and known region ($^{-}$145 to $^{-}$118 bp) containing asymetrically distributed purines and pyrimidines. Sequences containing alternating purine-pyrimidine and having the potential to take the Z-DNA conformation could also contribute to transcriptional control [32,33]; such sequences are found at $^{-}$369, $^{-}$229 and $^{-}$170 bp in rPOMC.

Transcription of the rPOMC gene is inhibited by glucocorticoids and stimulated by CRF and cyclic AMP ([13]; Gagner and Drouin, submitted). Glucocorticoid action is thought to be mediated by binding of the steroid to the glucocorticoid receptor protein. If the hormone-receptor complex acts directly at the DNA level to inhibit transcription, it could bind to sites homologous to those described in mouse mammary tumor virus (MMTV) [34,35]. The rPOMC 5′-flanking and exon 1 sequences do not contain the hexanucleotide sequence TGTYCT found in previously described glucocorticoid receptor binding sites [34–36]; however, many (11) homologous sequences are present. Some are found in exon 1 and in near upstream sequences: receptor binding in this region could inhibit transcription because of competition at overlapping binding sites for transcriptional factor(s). One hexanucleotide sequence ($^{-}$589 bp) has more extensive homology with a consensus derived by comparison of binding sites in MMTV, human metallothionein IIA and rabbit uteroglobin [36]. Experiments are underway to test for glucocorticoid receptor binding in this region and for glucocorticoid responsiveness in gene transfer experiments.

ACKNOWLEDGEMENTS

We are grateful to Guy Page, Barbara Cordell, Don Diamond and Keith Yamamoto for Long-Evans, NEDH, GH_3 and HTC genomic DNA, respectively. We thank Mamdouh Mikhail for help in computer sequence analysis and Diane Lapointe for preparation of this manuscript.

REFERENCES

[1] Eipper, B. and Mains, R. (1982) Endocr. Rev. 1, 1–27.
[2] Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S. (1979) Nature 278, 423–427.
[3] Civelli, O., Birnberg, N. and Herbert, E. (1982) J. Biol. Chem. 257, 6783–6787.
[4] Gee, C.E., Chen, C.-L.C. and Roberts, J.L. (1983) Nature 306, 374–376.
[5] Chen, C.-L.C., Mather, J.P., Morris, P.L. and Bardin, W. (1984) Proc. Natl. Acad. Sci. USA 81, 5672–5675.
[6] Chen, C.-L.C. and Chang, C.-C. (1985) Proc. 67th Annual Meeting Endocrine Society, 161.
[7] De Bold, C.R., Schworer, M.E., Connor, T.B., Bird, R.E. and Orth, D.N. (1983) Science 220, 721–723.
[8] Vale, W., Vaughan, J., Smith, M., Yamamoto, G., Rivier, J. and Rivier, C. (1983) Endocrinology 113, 1121–1131.
[9] Birnberg, N.C., Lissitzky, J.C., Hinman, M. and Herbert, E. (1983) Proc. Natl. Acad. Sci. USA 80, 6982–6986.
[10] Eberwine, J.H. and Roberts, J.L. (1984) J. Biol. Chem. 259, 2166–2170.

[11] Gagner, J.P. and Drouin, J. (1985) Mol. Cell. Endocrinol. 40, 25–32.
[12] Drouin, J. and Goodman, H.M. (1980) Nature 288, 610–613.
[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular CLoning, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[14] Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Natl. Acad. Sci. USA 77, 5172–5176.
[15] Nemer, M., Chamberland, M., Sirois, D., Argentin, S., Drouin, J., Dixon, R.A., Zivin, R.A. and Condra, J. (1984) Nature 312, 654–656.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5469.
[17] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[18] Takahashi, H., Teranishi, Y., Nakanishi, S. and Numa, S. (1981) FEBS Lett. 135, 97–102.
[19] Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) Nature 297, 335–339.
[20] Whitfeld, P.L., Seeburg, P.H. and Shine, S. (1982) DNA 1, 133–143.
[21] Takahashi, H., Hakamata, Y., Watanabe, Y., Kikuno, R., Miyata, T. and Numa, S. (1983) Nucleic Acids Res. 11, 6847–6857.
[22] Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429–438.
[23] Notake, M., Tobimatsu, T., Watanabe, Y., Takahashi, H., Mishina, M. and Numa, S. (1983) FEBS Lett. 156, 67–71.
[24] Uhler, M., Herbert, E., D'Eustachio, P. and Ruddle, D. (1983) J. Biol. Chem. 258, 9444–9453.
[25] Oates, E. and Herbert, E. (1984) J. Biol. Chem. 259, 7421–7425.
[26] Chambon, P., Dierich, A., Gaub, M.P., Jakowlev, S., Jongstra, J., Krust, A., LePennec, J.P., Oudet, P. and Reudelhuber, T. (1984) Recent Prog. Horm. Res. 40, 1–42.
[27] Weiher, H., Konig, M. and Gruss, P. (1983) Science 219, 626–631.
[28] Hearing, P. and Shenk, T. (1983) Cell 33, 695–703.
[29] Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4867–4871.
[30] Page, G.S., Smith, S. and Goodman, G.M. (1981) Nucleic Acid Res. 9, 2087–2104.
[31] Gidoni, D., Dynan, W.S. and Tijan, R. (1984) Nature 312, 409–413.
[32] Nordheim, A. and Rich, A. (1983) Nature 303, 674–679.
[33] Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M. and Beato, M. (1984) Nature 308, 513–519.
[34] Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.-A. and Yamamoto, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 6628–6632.
[35] Scheidereit, C., Geisse, S., Westphal, H.M. and Beato, M. (1983) Nature 304, 749–752.
[36] Cato, A.C.B., Geisse, S., Wenz, M., Westphal, H.M. and Beato, M. (1984) EMBO J. 3, 2771–2778.