Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6136–6140, September 1987
Biochemistry

Structure of the rat ornithine carbamoyltransferase gene, a large, X chromosome-linked gene with an atypical promoter

(urea cycle/mitochondrial enzyme/DNA sequence/sequence homology)

MASAKI TAKIGUCHI*, TAKASHI MURAKAMI*, SATOSHI MIURA†‡, AND MASATAKA MORI*§

*Institute for Medical Genetics, Kumamoto University Medical School, Kumamoto 862, Japan; and †Department of Biochemistry, Chiba University School of Medicine, Chiba 280, Japan

Communicated by Philip P. Cohen, May 26, 1987 (received for review April 24, 1987)

ABSTRACT **Rat mitochondrial ornithine carbamoyltransferase (EC 2.1.3.3) is encoded by a gene located on the X chromosome and expressed specifically in the liver and small intestine; we have cloned this gene and determined its structure. The gene is 75 kilobases long and is split into 10 exons. The introns range in length from 85 bases to 26 kilobases. The sum of the total exons is 1.5 kilobases and occupies only 2% of the gene; this value being one of the lowest among genes heretofore reported. The first exon encodes most of the NH_2-terminal presequence that functions as a mitochondrial targeting signal. Putative binding sites for the two substrates of the enzyme, carbamoyl phosphate and ornithine, are encoded by exons 3 and 9, respectively. A set of "CAAT box"- and "ATA box"-like sequences is present about 200 bases upstream from the 5′ end of the mRNA. About 35 bases downstream from this set of putative promoter elements, an 11-nucleotide sequence around the 5′ end of the mRNA reappears, as a direct repeat. This pair of direct repeats may play a role in pulling the cap site and the promoter elements together. Upstream and downstream from the 5′ end of the mRNA there are several sequences that resemble the transcription factor Sp1 binding site, the enhancer core sequence, the consensus sequence for the glucocorticoid receptor binding sites, and the putative enhancer element of the antithrombin III gene, another gene that is expressed specifically in the liver.**

Ornithine carbamoyltransferase (OrnCbmTase; carbamoylphosphate:L-ornithine carbamoyltransferase, EC 2.1.3.3) is a mitochondrial matrix enzyme encoded by a nuclear gene located on the X chromosome (1) and catalyzes the second step of urea synthesis in the liver of ureotelic animals. The enzyme is initially synthesized in the cytosol as a larger precursor with an NH_2-terminal presequence and is then transported into the mitochondrial matrix, where it is processed to the mature enzyme, concomitantly with the transport (refs. 2–4, reviewed in ref. 5). OrnCbmTase deficiency is known in humans (6) and mice (7, 8), as an inborn error of metabolism that results in protein intolerance and ammonia intoxication. The enzyme has been purified to homogeneity from bovine (9), rat (10), and human (11) liver and from rat small intestine (12). The primary structure of the enzyme precursor was deduced from the nucleotide sequences of rat (13–15) and human (16) cloned cDNAs.

The expression of the OrnCbmTase gene in mammals has several characteristics. The enzyme is found almost exclusively in the liver and, to a lesser extent, in the small intestine (12, 17). In the fetal liver, the level of the enzyme increases late in gestation, coordinately with other urea-cycle enzymes (18). The coordinate induction of the urea-cycle enzymes is caused also by an increase in dietary protein (19, 20). In addition to these tissue-specific, developmental stage-specific, and dietary regulations, the OrnCbmTase gene on one of the two X chromosomes in female cells is subject to inactivation, which accompanies X chromosome condensation (21, 22). Cloning and structure analysis of the OrnCbmTase gene will facilitate studies on molecular mechanisms of these modes of regulation of gene expression. Partial structures of the mouse (23) and human (24) OrnCbmTase gene have been reported. We have now cloned the rat OrnCbmTase gene and determined its complete structure. The gene is 75 kilobases (kb) long and consists of 10 exons. There are noteworthy structural elements around the promoter region that might participate in regulation of transcription.¶

MATERIALS AND METHODS

Isolation and Characterization of Genomic Clones. Three rat genomic libraries were used. Two libraries were prepared from liver DNA of male Sprague–Dawley rats by partial digestion with restriction endonuclease *Eco*RI (25) or *Hae* III respectively, using bacteriophage λ Charon 4A as a vector. These libraries were provided by T. D. Sargent, L. L. Jagodzinski, and J. Bonner (California Institute of Technology). The third library was prepared as follows. A partial *Sau*3A1 digest of liver DNA from a female Wistar rat was treated with calf intestine phosphatase and ligated with a *Bam*HI digest of bacteriophage λ EMBL4. After packaging *in vitro*, recombinant phages that carried rat DNA inserts were selected by use of a restrictive host strain, *Escherichia coli* NM539. Starting from 1 μg each of donor and vector DNAs, this procedure gave a library composed of 1.5×10^6 recombinants with an average insert size of 15 kb. These libraries were screened first with nick-translated cDNA fragments excised from the plasmid pOTC-1 (13) and later with labeled DNA fragments derived from some of the isolated genomic clones and free of repetitive sequences. Phage DNAs of positive clones were characterized by restriction mapping and by Southern (26) blot hybridization analysis. Subclones were constructed with plasmid pUC18 as a vector, and the nucleotide sequences of the double-stranded plasmid subclones were determined by the dideoxynucleotide chain-termination method (27, 28), using synthetic oligonucleotide primers complementary to the vector or cDNA sequences.

Analysis of the 5′ End of OrnCbmTase mRNA. Poly(A)$^+$ RNA was prepared from the liver and small intestine of a Wistar rat by guanidinium thiocyanate extraction (29) followed by oligo(dT)-cellulose chromatography. Determination of the 5′ end of the OrnCbmTase mRNA by nuclease S1

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: OrnCbmTase, ornithine carbamoyltransferase.

‡Present address: Meiji Institute of Health Science, 540 Naruda, Odawara, Kanagawa 250, Japan.
§To whom reprint requests should be addressed.

¶The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02957).

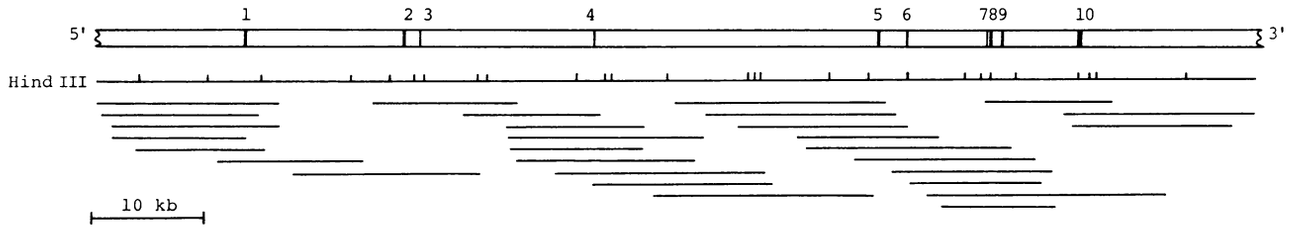

FIG. 1. Physical map of the rat OrnCbmTase gene. The structure of the gene is shown as the bar at the top of the diagram. Exons 1–10 are shown as vertical lines. Below the gene structure, *Hin*dIII sites are shown. The genomic DNA fragments contained in the phage clones are shown below the restriction map.

mapping (30) and primer extension (31) were performed as described.

RESULTS

Isolation and Characterization of the OrnCbmTase Gene. Three phage libraries constructed from rat liver genomic DNA were screened for the OrnCbmTase gene. About 40 independent clones were isolated and analyzed by restriction enzyme digestion. Representative clones are shown in Fig. 1. These clones overlapped and spanned about 100 kb. To define positions and boundaries of the exon blocks, the restriction fragments identified by Southern hybridization were subcloned and their sequences were determined (Figs. 1 and 2). The gene is 75 kb long and is divided into 10 exons. The nine introns range in size from 85 bases (intron 7) to 26 kb (intron 4). Since the exons total 1507 bases, 98% of the gene is occupied by the introns. All of the splice donor and acceptor sites conform to the GT . . . AG rule (32) for nucleotides immediately flanking exon borders.

The first exon encodes most (25 out of 32 amino acid residues) of the NH_2-terminal presequence that functions as a mitochondrial targeting signal. The presequence of the rat OrnCbmTase precursor was shown to direct a fused cytosolic protein to the mitochondria (33). Putative binding sites for the two OrnCbmTase substrates, carbamoyl phosphate and ornithine, are encoded by exons 3 and 9, respectively. Boxed arginine and cysteine residues in Fig. 2 were shown to be a part of the carbamoyl phosphate- and ornithine-binding sites on the basis of the following results. Inactivation of the enzyme by an arginine-reactive reagent, butanedione, is prevented by carbamoyl phosphate, and a cysteine-reactive reagent, 5,5′-dithiobis(2-nitrobenzoic acid), interferes with the binding of ornithine (34). Furthermore, these residues are conserved among the mammalian and *E. coli* OrnCbmTases (14).

Characterization of the 5′ and 3′ Ends of the OrnCbmTase Gene. DNA sequences around the 5′ and 3′ end of the OrnCbmTase gene are shown in Fig. 3. The 5′ end of the mRNA was determined by nuclease S1 mapping and primer extension (Fig. 4), and is numbered +1. The DNA fragment labeled at a position 21 bases downstream from the initiation codon was used as the probe for the S1 mapping and gave protected fragments of 106–108 bases for both liver and small intestinal poly(A)$^+$ RNA. The 108-base-long fragment, starting from the primer labeled at the same position as the S1 probe, was also detected by primer extension analysis. From these results, the 5′ end of the OrnCbmTase mRNA was assigned to a position 87 bases upstream from the initiation codon. The assigned 5′ end was the residue A preceded by C, the generally preferred cap site.

The sequences CAAT (−235 to −232) and ATAA (−191 to −188), resembling canonical "CAAT" and "ATA" boxes, are situated more upstream than the ordinary locations, which are around 80 and 30 bases upstream from a cap site, respectively

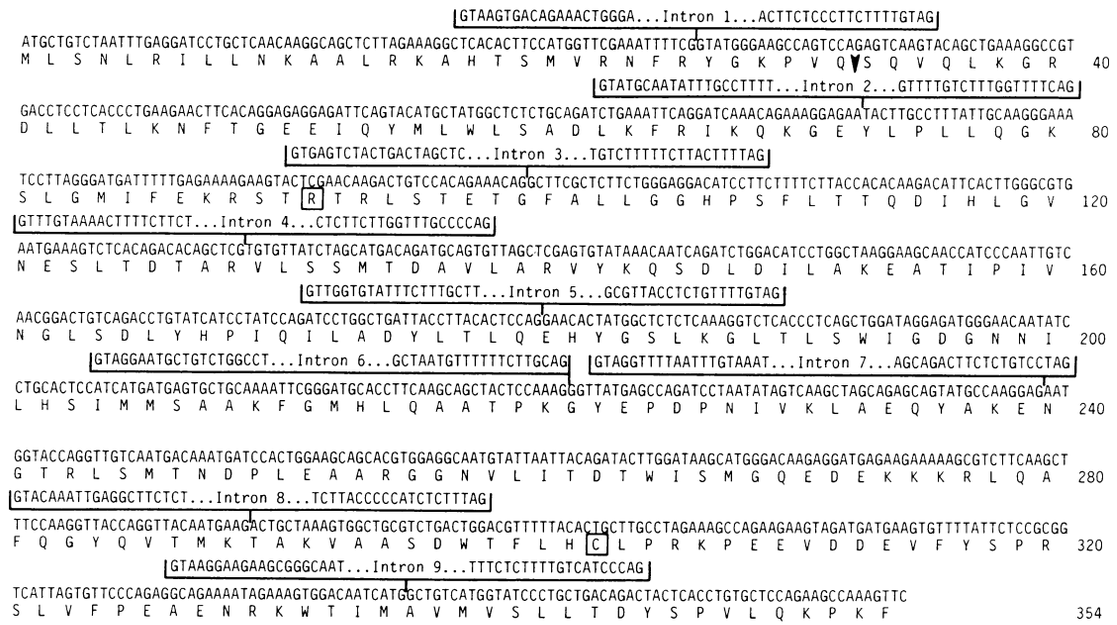

FIG. 2. Exon–intron organization of the rat OrnCbmTase gene. The locations of the nine introns are indicated above the contiguous protein-coding sequence. The amino acid sequence (standard one-letter symbols) is given below the coding sequence. 5′ and 3′ untranslated sequences are not included. Arrowhead indicates the position of proteolytic maturation. Boxed arginine and cysteine residues are the putative binding sites for the substrates, carbamoyl phosphate and ornithine, respectively.

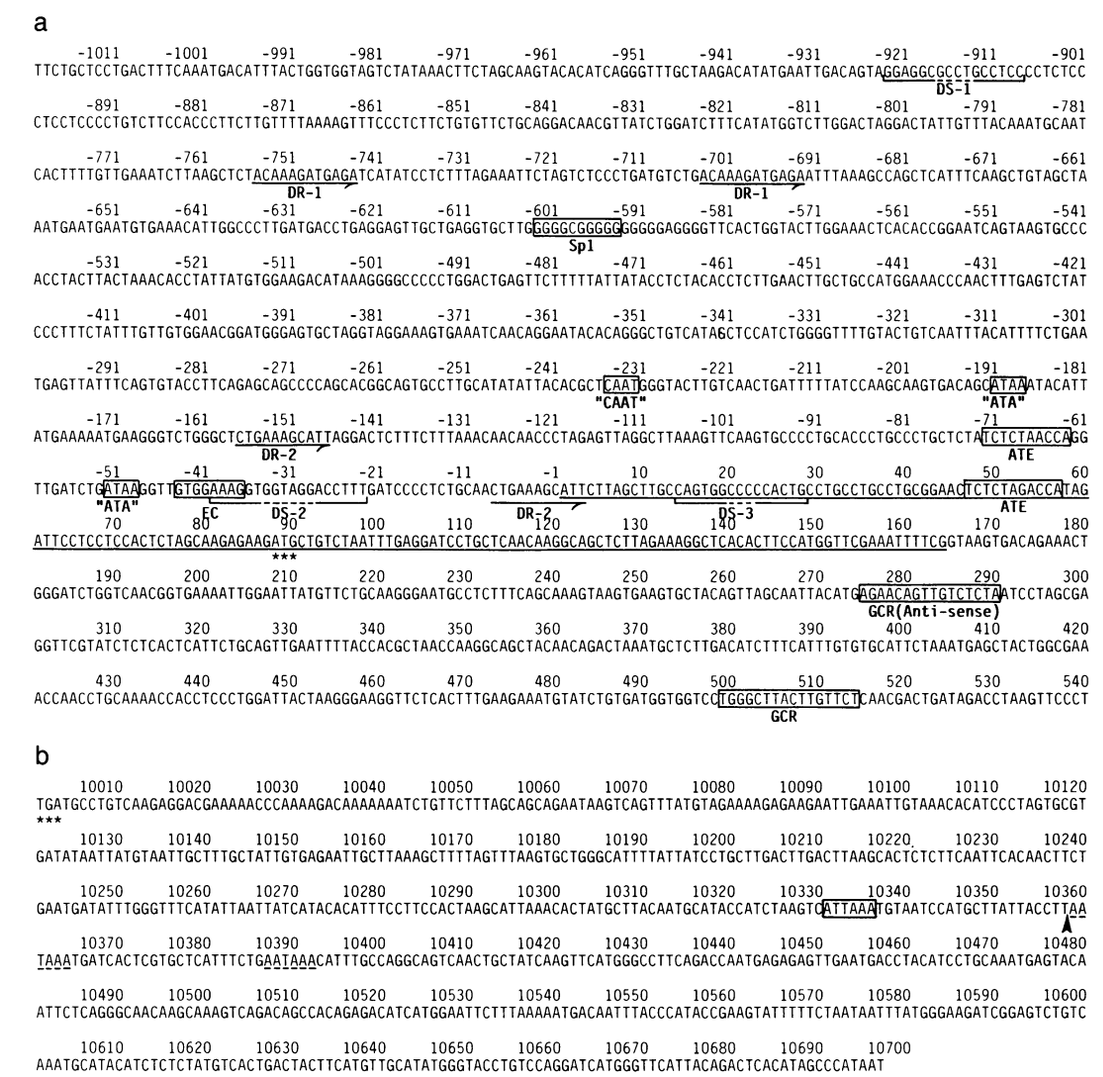

FIG. 3. Nucleotide sequences of 5′ (*a*) and 3′ (*b*) termini. The sequence was determined on both strands, except for the regions upstream from position −997 and downstream from +492 in the 5′ terminus. (*a*) The underline (+1 to +164) denotes the first exon. The initiation codon is marked with stars. The boxed areas with the symbols resemble the following sequences; "ATA", ATA box; "CAAT", CAAT box; Sp1, Sp1-binding site (35); EC, enhancer core sequence (36); GCR, consensus sequence for glucocorticoid receptor binding sites (37); ATE, putative enhancer element of human antithrombin III gene (38). The brackets with the intervening broken lines, marked with DS-1 to -3, indicate the sequences of dyad symmetry. Paired arrows DR-1 and DR-2 indicate direct repeats. (*b*) The sequence downstream of the termination codon TGA (marked with stars) is shown. The nucleotide corresponding to the termination codon is numbered 10,001. Arrowhead indicates the poly(A) addition site. The putative poly(A)-addition signal is boxed. The "typical" poly(A)-addition signal sequences are denoted by the broken lines.

(Fig. 3*a*). About 35 bases downstream from the ATAA sequence is located the 11-nucleotide sequence CTGAAAGCATT (−157 to −147), this being a direct repeat of the sequence around the 5′ end of the mRNA (−8 to +3). Therefore, this alignment of CAAT, ATAA, and the cap site-like region forms a typical structure that could function as a promoter, although no transcript starting from this region was detected. Another ATAA sequence is present at positions −52 to −49, and even this position is more upstream than the usual one.

A potential binding site for the cellular transcription factor Sp1 (35), GGGGCGGGGG, is found at positions −603 to −594. A sequence similar to the viral and cellular enhancer core sequence (36), GTGGAAAG, is situated at positions −44 to −37. The latter half of this enhancer core-like sequence is overlapped by a sequence of dyad symmetry, AAAGGTggtaggACCTTT (−40 to −23, uppercase letters show the complementary regions). Two sequences in the first intron, located at positions +290 to +275 (antisense strand) and +499 to +514, resemble the consensus sequence for the glucocorticoid receptor binding sites (37) (Fig. 5*a*). Two sequences surrounding the 5′ end of the mRNA, located at positions −72 to −63 and +47 to +57, are similar to the sequence that was postulated as an enhancer element of the human antithrombin III gene (38), expressed mainly in the liver, which in turn resembles a portion of the polyoma virus enhancer (39) (Fig. 5*b*).

The 3′ untranslated region of the OrnCbmTase gene contains 358 nucleotides (Fig. 3*b*). The site of polyadenylylation was inferred from three independent cDNA clones (refs. 14 and 15 and unpublished results). An uncommon but not unprecedented poly(A)-addition signal, ATTAAA (40, 41), is located 22 bases upstream from the poly(A) tail. The typical poly(A)-addition signal sequence AATAAA appears twice, just following the polyadenylylation site and 28 bases downstream from the site. Whether or not these sequences function as minor poly(A)-addition signals is unknown.

DISCUSSION

The present study shows that the rat OrnCbmTase gene spans about 75 kb on the X chromosome. Since the sum of the

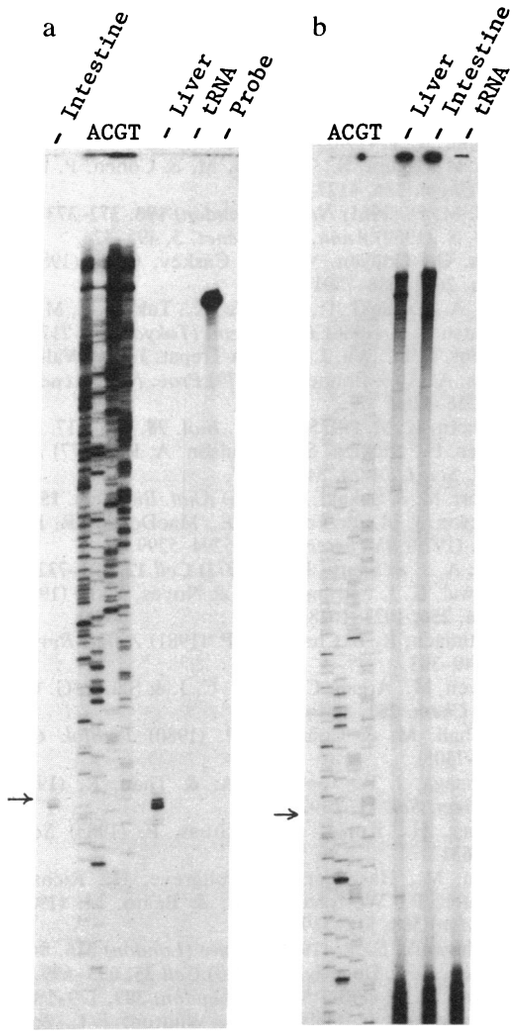

FIG. 4. Determination of the 5′ end of rat OrnCbmTase mRNA by nuclease S1 mapping (*a*) and primer extension (*b*). Poly(A)$^+$ RNAs from rat liver and small intestine were analyzed. (*a*) The probe was an *Apa* I–*Bam*HI fragment (−500 to +108) ^{32}P-labeled at the 5′ end of the *Bam*HI site. Hybridization was carried out with 0.02 pmol of probe (1.3×10^5 dpm) and 5 μg of poly(A)$^+$ RNA at 51°C for 20 hr, and the mixture was subjected to nuclease S1 digestion. (*b*) The primer was a synthetic oligonucleotide (20-mer, complementary to positions +89 to +108) ^{32}P-labeled at the 5′ end. Annealing was carried out with 0.3 pmol of primer (4×10^6 dpm) and 5 μg of poly(A)$^+$ RNA, and the mixture was subjected to reverse transcription. In both *a* and *b*, yeast tRNA was used as a control. To compare the products directly with the genomic nucleotide sequence, the 5′ ends of the probe and the primer were arranged to coincide with the 5′ end of a primer of the dideoxynucleotide sequencing (reaction mixtures A, C, G, and T). Portions of samples were electrophoresed in a 6% acrylamide/7 M urea gel. Arrow indicates the position assigned to the 5′ end of the mRNA.

exon lengths is 1.5 kb, exons occupy only 2% of the whole gene, a value being one of the lowest among genes heretofore reported. X chromosome-linked genes whose structures have been determined thus far are relatively large: the hypoxanthine phosphoribosyltransferase gene spans 33 kb (42); the factor IX gene, 34 kb (43); the factor VIII gene, 186 kb (44); the phosphoglycerate kinase gene, 23 kb (45); the color-vision genes, 13–15 kb (46); and the glucose-6-phosphate dehydrogenase gene, 18 kb (47). Though the number of examples is small and they were not selected by random sampling, X chromosome-linked genes generally appear to be large. It is tempting to speculate that large genes are more favorable for inactivation associated with X chromosome condensation.

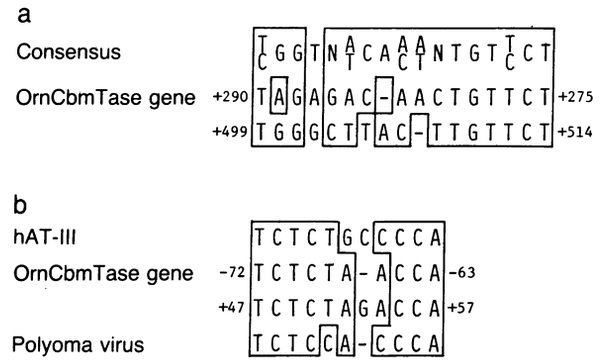

FIG. 5. Similarities between the 5′-region sequences of the OrnCbmTase gene and the consensus sequence for the glucocorticoid receptor binding sites (37) (*a*) and the putative enhancer element of the human antithrombin III gene (hAT-III, ref. 38) (*b*), which in turn resembles a portion of the polyoma virus enhancer (39). The OrnCbmTase gene sequence at position +290 to +275 is on the antisense strand. Matched nucleotides are boxed. Dashes represent gaps introduced for maximal sequence identity.

The 5′ end of the rat OrnCbmTase mRNA assigned here is 87 bases upstream from the initiation codon, both in the liver and the small intestine. Among the three OrnCbmTase cDNA clones reported, one clone (14) starts just from the cap site, another from a position 30 bases downstream (13), and the third from a position 13 bases upstream (15). The last one was probably derived from mRNA that initiated from the upstream, less frequent start site.

The sequences CAAT and ATAA are present 235 bases and 191 bases upstream from the 5′ end of the mRNA. It remains open to question whether these sequences function as "CAAT" and "ATA" boxes, since their positions are far upstream from the usual locations. However, it is noteworthy that the 11-nucleotide sequence CTGAAAGCATT around the 5′ end of the mRNA is directly repeated about 35 bases downstream from the sequence ATAA. Therefore, it is possible that the cap site and the ATAA sequence are pulled into the same vicinity by interaction of the direct repeats with their binding protein(s) and by the resultant "looping out" of the interval sequence (48). In this manner, the putative CAAT and ATA boxes may become functional.

The structures of the 5′ regions of the mouse (23) and human (24) OrnCbmTase genes have been reported. The entire reported sequence of the mouse gene (corresponding to positions −636 to +189 of the rat gene) is 92% identical with the rat sequence except for several insertions or deletions, and that of the human gene (corresponding to positions −500 to +305) is 75% identical with the rat sequence (data not shown). These results indicate that rather long areas of the 5′ region of the gene, including the first intron, are highly conserved among these species. For the mouse OrnCbmTase gene, an ≈800-nucleotide 5′ flanking region was reported (23) to direct the expression of the fused chloramphenicol acetyltransferase gene in human hepatoma cells.

Around the 5′ end of the mRNA are noteworthy sequences resembling the previously characterized cis-acting transcriptional elements (Fig. 3*a*). The sequence GGGGCGGGGG (−603 to −594) contains the core GGGCGG sequence of the binding site of transcription factor Sp1 (35), which enhances transcription by RNA polymerase II, and coincides with one of the Sp1-binding sequences of the immediate early gene *3* of herpes simplex virus 1 (35). The homologue of the enhancer core sequence (36), GTGGAAAG (−44 to −37), is overlapped by a sequence of dyad symmetry that has the potential to serve as a recognition site of DNA-binding protein(s) (49); therefore, this site might be bound competitively by various proteins. The sequences located in the first

intron at positions +290 to +275 (antisense strand) and +499 to +514, which resemble the consensus sequence for the glucocorticoid receptor binding sites (37) (Fig. 5*a*), might be involved in induction of OrnCbmTase activity by the cooperation of dexamethasone and glucagon (50, 51). As shown in Fig. 5*b*, two sequences near the 5′ end of the mRNA are similar to the sequence of a putative enhancer element of the antithrombin III gene (38) expressed mainly in the liver. Therefore, the homologous sequences of the OrnCbmTase gene might also function as liver-specific enhancers. These characteristic sequences remain to be tested for interaction with protein factors and for their effects on transcription of the OrnCbmTase gene.

We thank Dr. T. D. Sargent for the genomic libraries, Dr. T. Hayashi of the National Cancer Research Institute for introducing us to screening of the libraries, Drs. F. Kishi and A. Nakazawa of Yamaguchi University for the vector EMBL4 and the strain NM539, Dr. Y. Ebina of our laboratory (Kumamoto University) for valuable discussions, and Ms. M. Ohara of Kyushu University for comments on the manuscript. This work was supported in part by Grants-in-Aid 61480123, 61870019, and 61770185 from the Ministry of Education, Science, and Culture of Japan and a grant from the Yamada Science Foundation.

1. Short, E. M., Conn, H. O., Snodgrass, P. J., Campbell, A. G. M. & Rosenberg, L. E. (1973) *N. Engl. J. Med.* **288,** 7–12.
2. Conboy, J. G., Kalousek, F. & Rosenberg, L. E. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5724–5727.
3. Mori, M., Miura, S., Tatibana, M. & Cohen, P. P. (1980) *J. Biochem. (Tokyo)* **88,** 1829–1836.
4. Mori, M., Morita, T., Ikeda, F., Amaya, Y., Tatibana, M. & Cohen, P. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6056–6060.
5. Mori, M., Miura, S., Morita, T., Takiguchi, M. & Tatibana, M. (1982) *Mol. Cell. Biochem.* **49,** 97–111.
6. Walser, M. (1983) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), 5th Ed., pp. 402–419.
7. DeMars, R., LeVan, S. L., Trend, B. L. & Russell, L. B. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 1693–1697.
8. Qureshi, I. A., Letarte, J. & Ouellet, R. (1979) *Pediatr. Res.* **13,** 807–811.
9. Marshall, M. & Cohen, P. P. (1972) *J. Biol. Chem.* **247,** 1641–1653.
10. Lusty, C. J., Jilka, R. L. & Nietsch, E. H. (1979) *J. Biol. Chem.* **254,** 10030–10036.
11. Pierson, D. L., Cox, S. L. & Gilbert, B. E. (1977) *J. Biol. Chem.* **252,** 6464–6469.
12. Wraight, C., Lingelbach, K. & Hoogenraad, N. (1985) *Eur. J. Biochem.* **153,** 239–242.
13. Takiguchi, M., Miura, S., Mori, M., Tatibana, M., Nagata, S. & Kaziro, Y. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7412–7416.
14. Kraus, J. P., Hodges, P. E., Williamson, C. L., Horwich, A. L., Kalousek, F., Williams, K. R. & Rosenberg, L. E. (1985) *Nucleic Acids Res.* **13,** 943–952.
15. McIntyre, P., Graf, L., Mercer, J. F. B., Wake, S. A., Hudson, P. & Hoogenraad, N. (1985) *DNA* **4,** 147–156.
16. Horwich, A. L., Fenton, W. A., Williams, K. R., Kalousek, F., Kraus, J. P., Doolittle, R. F., Konigsberg, W. & Rosenberg, L. E. (1984) *Science* **224,** 1068–1074.
17. Ryall, J., Nguyen, M., Bendayan, M. & Shore, G. C. (1985) *Eur. J. Biochem.* **152,** 287–292.
18. Lamers, W. H., Mooren, P. G., De Graaf, A. & Charles, R. (1985) *Eur. J. Biochem.* **146,** 475–480.
19. Schimke, R. T. (1962) *J. Biol. Chem.* **237,** 459–468.
20. Mori, M., Miura, S., Tatibana, M. & Cohen, P. P. (1981) *J. Biol. Chem.* **256,** 4127–4132.
21. Lyon, M. F. (1961) *Nature (London)* **190,** 372–373.
22. Ohno, S. (1969) *Annu. Rev. Genet.* **3,** 495–524.
23. Veres, G., Craigen, W. J. & Caskey, C. T. (1986) *J. Biol. Chem.* **261,** 7588–7591.
24. Hata, A., Tsuzuki, T., Shimada, K., Takiguchi, M., Mori, M. & Matsuda, I. (1986) *J. Biochem. (Tokyo)* **100,** 717–725.
25. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A. & Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3256–3260.
26. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238.
29. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
30. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12,** 721–732.
31. Agarwal, K. L., Brunstedt, J. & Noyes, B. E. (1981) *J. Biol. Chem.* **256,** 1023–1028.
32. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
33. Nguyen, M., Argan, C., Lusty, C. J. & Shore, G. C. (1986) *J. Biol. Chem.* **261,** 800–805.
34. Marshall, M. & Cohen, P. P. (1980) *J. Biol. Chem.* **255,** 7287–7305.
35. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23.
36. Weiher, H., Konig, M. & Gruss, P. (1983) *Science* **219,** 626–631.
37. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M. & Beato, M. (1984) *Nature (London)* **308,** 513–519.
38. Prochownik, E. V. (1985) *Nature (London)* **316,** 845–848.
39. Linney, E. & Donerly, S. (1983) *Cell* **35,** 693–699.
40. Berget, S. M. (1984) *Nature (London)* **309,** 179–182.
41. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13,** 1347–1368.
42. Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2147–2151.
43. Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) *EMBO J.* **3,** 1053–1060.
44. Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J. & Lawn, R. M. (1984) *Nature (London)* **312,** 326–330.
45. Michelson, A. M., Blake, C. C. F., Evans, S. T. & Orkin, S. H (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6965–6969.
46. Nathans, J., Thomas, D. & Hogness, D. S. (1986) *Science* **232,** 193–202.
47. Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M. & Persico, M. G. (1986) *EMBO J.* **5,** 1849–1855.
48. Ptashne, M. (1986) *Nature (London)* **322,** 697–701.
49. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321.
50. Gebhardt, R. & Mecke, D. (1979) *Eur. J. Biochem.* **97,** 29–35.
51. Lin, R. C., Snodgrass, P. J. & Rabier, D. (1982) *J. Biol. Chem.* **257,** 5061–5067.