Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4912–4916, July 1989
Biochemistry

Structural characterization of the rat malic enzyme gene

H. MORIOKA*, M. A. MAGNUSON†, T. MITSUHASHI*, M.-K. H. SONG*, J. E. RALL*, AND V. M. NIKODEM*‡

*Clinical Endocrinology Branch, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892; and †Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232

Contributed by J. E. Rall, April 3, 1989

ABSTRACT **We have identified and characterized λ bacteriophage clones containing genomic DNA encoding rat malic enzyme [(*S*)-malate:NADP$^+$ oxidoreductase (oxaloacetate-decarboxylating); EC 1.1.1.40]. The malic enzyme gene is unexpectedly large, spanning at least 95 kilobases. It is divided into 14 exons that range in size from 76 to 1513 base pairs. The sizes and boundaries of the exons were determined by Southern blotting and DNA sequencing. The sequences at the 5′ and 3′ ends of each intron conformed to the consensus sequence for mammalian introns. S1 nuclease and primer-extension assays showed that transcription of the malic enzyme gene initiates at multiple sites, the strongest one at position −31 relative to the ATG. "TATA and CCAAT box" homologies are not present in the proximal promoter region. Analysis of the 3′ end of the gene showed that the utilization of alternate polyadenylylation signals in exon 14 results in two mRNAs with 3′ untranslated regions of 345 and 1345 nucleotides, respectively.**

Cytosolic malic enzyme [ME; (*S*)-malate:NADP$^+$ oxidoreductase (oxaloacetate-decarboxylating; EC 1.1.1.40] catalyzes the NADP-dependent oxidative decarboxylation of malate to pyruvate and CO_2 generating NADPH. In liver the enzyme is under nutritional and hormonal regulation. Studies from this laboratory indicate that the dietary, thyroid hormone, and dehydroepiandrosterone regulation of rat hepatic ME mRNA occur by different mechanisms. The accumulation of ME mRNA after feeding a high-carbohydrate diet to euthyroid rats results solely from increased cytoplasmic stability of the ME mRNAs (1). The increase of ME gene expression in response to 3,5,3′-triiodo-L-thyronine (T3) is due to both transcriptional activation of the gene and stabilization of the ME pre-mRNA (2). Whereas the ≈8-fold increase in cytoplasmic ME mRNA in liver of rats treated with dehydroepiandrosterone can be fully accounted for by changes in the rate of transcription (unpublished results).

To understand the molecular mechanisms by which ME is regulated, to identify sequences in the ME gene that are involved in these regulations, and to disclose the structural basis for the generation of the two ME mRNAs, we have characterized genomic DNA encoding ME. In the present study, we report the isolation and characterization of cloned sequences representing the rat ME gene.§ By restriction endonuclease mapping and direct sequencing of exon–intron boundaries, we have found that this gene is unexpectedly large and contains at least 14 exons. Analysis of the transcription start sites by S1 nuclease and primer-extension analyses revealed multiple sites of transcription initiation. This is in accord with the absence of two frequently identified promoter elements, the so-called "TATA and CCAAT boxes," which are usually located in the vicinity of cap sites and are necessary for precise and efficient initiation of transcription (4, 5). Comparison of the genomic and the cDNA sequences downstream from the last exon in conjunction with S1 nuclease analysis of this region revealed that the two ME mRNAs arise from the use of alternative polyadenylylation signals.

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.

MATERIALS AND METHODS

Isolation and Characterization of ME Genomic Clones. Three rat genomic DNA libraries were used to isolate the gene for ME. The λ Charon 4A Libraries (C), made by partial *Eco*RI or *Hae* III digestion of rat genomic DNA, were obtained from J. Bonner (Phytogen, Pasadena, CA) or Clontech, respectively, and a λ EMBL 3 library (E) made by partial digestion with *Sau*3A was obtained from G. Scherer (Institute for Human Genetics, Freiberg, F.R.G.). The libraries were screened by the method of Benton and Davis (6) using various ME cDNA fragments as probes (7). The DNAs were ^{32}P-labeled by either nick-translation or random nucleotide priming (Pharmacia; oligo labeling kit) using [α-^{32}P]dCTP (>3000 Ci/mmol; 1 Ci = 37 GBq). To identify ME genomic clones that were not isolated using cDNA probes, we used several 42-mer oligonucleotides complementary to ME cDNA that were synthesized on a model 380A Applied Biosystems DNA synthesizer. The oligonucleotides were 5′-end-labeled using [γ-^{32}P]ATP (>5000 Ci/mmol) and polynucleotide kinase and then were hybridized to filters as described by Ullrich *et al.* (8).

Restriction maps of ME genomic DNA were determined by double restriction endonuclease digests of λ DNA and plasmid subclones containing genomic DNA fragments. Overlapping regions of genomic DNA deduced from common restriction endonuclease patterns were confirmed in some cases by cross-hybridization using genomic DNA fragments as probes. Exons were located by hybridizing ME cDNA fragments to Southern blots of restriction endonuclease digested cloned genomic DNA. The smallest restriction fragments that could be identified by this approach were gel-purified and subcloned into appropriate vectors for dideoxynucleotide sequence analysis (9). Intron–exon boundaries were assigned by comparing the genomic and the cDNA sequences.

Primer-Extension Analysis. A subclone of λ29 (pME29.8) was digested with *Eco*RI and *Ava* I; the fragment was treated with calf intestinal alkaline phosphatase and end-labeled using [γ-^{32}P]ATP and polynucleotide kinase. After inactivation of the kinase, the DNA was digested with *Bst*XI and the 5′ end-labeled 23-base-pair (bp) fragment was isolated from an 8% sequencing gel (10). Twenty-five micrograms of poly(A)$^+$ RNA, prepared from T3-treated rat liver (50 μg/100 g of body weight for 7 days) by the guanidinium thiocyanate method (11), was mixed with 2×10^5 cpm of the labeled fragment in the reaction mixture (12), denatured at 80°C for 15 min, and allowed to hybridize at 65°C overnight. To initiate primer-extension, 0.43 mM of each of the four deoxyribonucleotide triphosphates and reverse transcriptase at 54 units/

Abbreviations: T3, 3,5,3′-triiodo-L-thyronine; nt, nucleotide(s); ME, malic enzyme.

‡To whom reprint requests should be addressed.

§The λME36 sequence reported in this paper has been deposited in the GenBank data base (accession no. M12545).

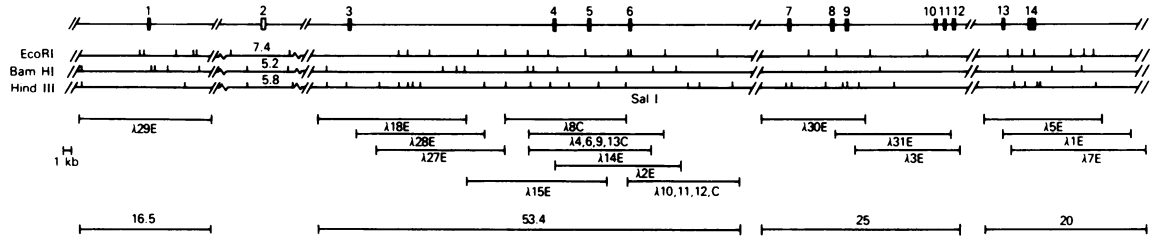

FIG. 1. Restriction map and structural organization of rat ME gene. The top line denotes the gene from 5′ to 3′ (left to right). The positions and sizes of the exons are drawn approximately to scale; the solid boxes show the positions of those exons that were mapped by sequencing, the open box marks the missing exon, and interrupted lines are missing intron sequences. An approximate composite restriction map for the indicated enzymes is shown below; the *Sal* I site is located at the 3′ end of exon 6. Genomic clones aligned according to the restriction mapping are designated by numbers and C and E determine the genomic library used. The bottom line depicts the length in kb of a continuous ME genomic DNA available.

ml were added and the mixture was incubated at 42°C for 1 hr. Extension products were analyzed on an 8% sequencing gel, followed by autoradiography.

S1 Nuclease Analysis. Nae *I*–Ava *I probe*. A subclone of λ29 (pME29.8) was digested with *Ava* I, end-labeled using [γ-^{32}P]ATP and polynucleotide kinase, then digested with *Nae* I. The 5′-end-labeled 223-bp fragment was gel-purified and then used to analyze the 5′ end of the ME gene.

Eco*RI*–Pst *I probe*. A subclone of λ1 (pME1.1) was digested with *Eco*RI, end-labeled using Klenow fragment of DNA polymerase I and [α-^{32}P]ATP, and then digested with *Pst* I. The 3′-end-labeled 1430-bp fragment was gel-purified and used to analyze the 3′ end of ME gene.

S1 analysis was performed as described (13), but the overnight hybridization temperature was 65°C for the analysis of the 5′ end and 43°C for the 3′ end of the gene. A typical reaction contained 25 μg of poly(A)$^+$ RNA prepared as described above. The hybridization mixtures were digested with 48 units of S1 nuclease at 37°C for 30 min and then analyzed by electrophoresis followed by autoradiography.

Computer Analysis of RNA Secondary Structure. The 5′ untranslated region of the ME gene was analyzed using the DNA:SEQH program (National Institutes of Health Computing Center) based on the method of Tinoco (14) to calculate ΔG values.

Materials. Restriction enzymes, Klenow fragment of *Escherichia coli* DNA polymerase I, T4 polynucleotide kinase, and avian myeloblastosis virus reverse transcriptase were obtained from New England Biolabs or Bethesda Research Laboratories and used as recommended by the suppliers. The labeled nucleotides were purchased from New England Nuclear.

RESULTS AND DISCUSSION

Cloning and Structural Organization of the Rat ME Gene. The genomic DNA fragments shown in Fig. 1 were obtained by screening λ genomic libraries with various rat ME cDNA sequences as probes (7). Seventeen genomic clones were studied further by restriction mapping and Southern blot analysis to localize specific exons. The genomic fragments containing exon sequences and their adjacent regions were subcloned into plasmid vectors and sequenced according to the strategy indicated in Fig. 2, to obtain the precise location of each of the 5′ and 3′ exon–intron boundaries. Comparison of the genomic DNA sequences with the cDNA sequences (7) enabled us to determine the structural organization of the rat ME gene (Fig. 1). The sequences of exon–intron junctions and the size of exons and introns are summarized in Table 1. All splice junction sequences adhere to the GT/AG rule

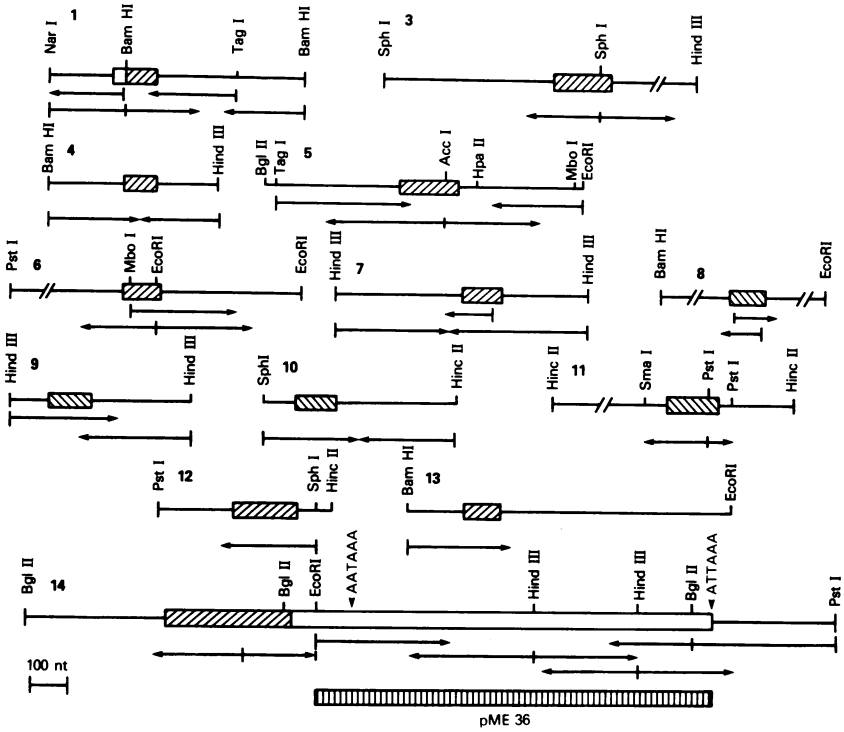

FIG. 2. Sequence strategy of determining the exon–intron junctions and a 3′ untranslated region of the ME gene. The appropriate subcloned genomic fragments in pUC or M13 vectors derived from λ recombinants were used for DNA sequencing. The arrows indicate the direction of sequencing, the hatched bars are the coding sequences, the solid lines are intronic regions, and the open bars are untranslated regions. The location of ME36, a cDNA clone containing the 3′ untranslated region of the longer ME mRNA is indicated below exon 14. The shorter mRNA terminates 24 nt after the AATAAA; the longer mRNA terminates 12 nt after the ATTAAA.

Table 1. Exon–intron arrangements of rat ME gene and sequences at the splice junctions

Exon	Exon size, bp	3′ exon junction	Intron	5′ exon junction	Exon
1	109*	26		27	
		Asn Lys		Asp Leu	
		AAC AAG	gtgagc———(A > 7.5 kb)———	GAC TTG	2
2	134	71		71	
		Asp Ar		g Tyr	
		GAC AG	———(B > 5 kb)———tccacag	G TAT	3
3	150	121		121	
		Pro Ar		g Gly	
		CCA AG	gtataa———(C > 24 kb)———ctaccag	A GGC	4
4	76	146		147	
		Val Lys		Ala Ile	
		GTC AAG	gtaagt———(D ≈ 4.5 kb)——cttgtag	GCT ATT	5
5	162	200		201	
		Asn Glu		Glu Leu	
		AAT GAG	gtaaac———(E ≈ 5 kb)———ccactag	GAG TTA	6
6	104	235		235	
		Ser Ly		s Tyr	
		TCC AA	gtatag———(F > 15 kb)———ctcatag	A TAT	7
7	110	272		272	
		Gln G		ly Thr	
		CAA G	gtaatc———(G ≈ 5 kb)————ttttag	GA ACA	8
8	98	304		305	
		Gly Glu		Ala Ala	
		GGC GAG	gtatct———(H ≈ 1 kb)————tttcag	GCT GCC	9
9	114	342		343	
		Val Lys		Gly Arg	
		GTT AAG	gtaaga———(I ≈ 11 kb)———ttcttag	GGG CGT	10
10	106	378		378	
		Ile G		ly Val	
		ATA G	gtgcttt———(J ≈ 0.84 kb)——ttcatag	GA GTT	11
11	143	425		426	
		Thr Lys		Gly Arg	
		ACC AAG	gtgagc———(K ≈ 1.4 kb)——gttgcag	GGC CGT	12
12	174	483		484	
		Ala Glu		Val Ile	
		GCTCTAG	gtgact———(L > 9.5 kb)———tttaag	GTC ATA	13
13	99	516		517	
		Val Lys		Ile Val	
		GTA AAG	gtagcat———(M ≈ 3.3 kb)——tttttag	ATT GTG	14
14(S)	513	GATTAC			
14(L)	1513	GAGATA			

Sequences at the splice junctions were determined by comparing the genomic DNA sequence with the cDNA sequence. Exon sequences are in capital letters; intron sequences are in lower case letters. The amino acid residues are numbered (7). The exon 14 containing 345 nt of untranslated region is designated 14(S) and that of 1345 nt 14(L). The size of the introns was estimated by restriction enzyme mapping of the λ clones (Fig. 1).
*Determined from position +1, shown in Fig. 4.

(4)—i.e., they begin at the 5′ end with the dinucleotide GT and terminate with the dinucleotide AG. A total of 14 exons and 13 introns were found (Fig. 1 and Table 1), when 115 kilobases (kb) of genomic DNA was analyzed in this fashion.

All exons, with the exception of the 3′ exon [513 nucleotides (nt) for the short mRNA or 1513 nt for the longer mRNA], show a restricted size distribution typical for eukaryotic genes (15), ranging from 76 to 174 bp. Three regions of the ME gene were not represented in the genomic DNA isolated. One region contains the second exon and surrounding introns. The exon was mapped by Southern blot analysis of total genomic DNA and appears to be contained in a 5.2-kb *Bam*HI fragment (data not shown). It is unlikely that this region could encompass two exons because of the absence of an additional AG/G consensus splicing sequence. Consequently, the size of introns A and B (Table 1) cannot be determined but they are larger than 7.5 and 5 kb, respectively. The size of the other two introns, F and L, is incomplete because the genomic clones containing exon 6 and 7 and clones containing exons 12 and 13 do not overlap in their restriction enzymes sites, but they are larger than 15 and 9.5 kb, respectively.

We screened more than 5×10^6 recombinants from three genomic libraries and did not identify these missing regions. In some cases, the filters used for the screening of these missing portions of the ME gene were subsequently used successfully for obtaining genomic DNA containing other exons or introns. It is likely that these missing sequences are underrepresented in the genomic libraries used. Since three restriction enzymes were used for the construction of these libraries, two enzymes at limited digestion of their 4-bp recognition sites, the possibility of selection against these genomic sequences during cloning is unlikely. It seems more likely that the poor growth of clones containing these sequences accounts for our inability to identify these sequences. Bass *et al.* (16) reported similar difficulties when isolating the human thyroglobulin gene. This is supported by the observation of Wyman *et al.* (17) that 9% of the recom-

binant phage in λ Charon 30 fail to grow in an *E. coli* host, whereas they grow well in RecB and RecC hosts.

Thus, the rat ME gene is more than 95 kb long and contains several introns larger than 15 kb. Large genes have been reported, factor VIII (18) or thyroglobulin (16); however, the rat ME enzyme gene is the largest of the housekeeping genes so far characterized (19). This large size for the ME gene could possibly explain a delay in the accumulation of the ME mRNA as a response to thyroid hormone stimulation. Towle *et al.* (20) and our own studies have failed to detect an increase in ME mRNA after T3 treatment of rats until about 2 hr. This had led some to speculate that the induction of ME by T3 might occur by a secondary event. With an estimated transcriptional rate of 2000 nt per min (21), a minimum of an hour would be required for RNA polymerase to copy the entire gene. Thus a reason for the delayed response of ME to T3 treatment may simply be the time required for transcription of the gene.

The function and significance of large introns in genes are not clear. It is possible that other genes are present within these large introns and code for different cytoplasmic RNAs (22). Another hormonally regulated gene expressed in the liver is that for phospho*enol*pyruvate carboxykinase, which encodes a 2624-nt mRNA in a gene of only 6 kb (23). ME, in comparison, encodes mRNAs of approximately 2100 and 3100 nt in a gene of at least 100 kb. Thus the ratio of mRNA size to gene size varies from about 1:2 for phospho*enol*pyruvate carboxykinase to approximately 1:50 for ME. These differences might affect the rate of response of these genes to hormones and other agents.

Characterization of the 5′ End of the ME mRNAs. The 5′ termini of the ME mRNAs were identified by S1 nuclease mapping and primer-extension analysis. For primer-extension analysis, a 5′-^{32}P-labeled 23-bp *Bst*XI–*Ava* I fragment from λ29 (Fig. 3*C*) was hybridized to rat liver poly(A)$^+$ RNA. This hybrid was then used to prime DNA synthesis by reverse transcriptase. The elongated products were run on sequencing gels alongside a sequencing reaction for a cDNA fragment. Fig. 3*A* shows that the size of the most abundant extended products corresponds to 5′ untranslated regions 31 and 30 nt long with respect to the adenine of the translation initiation codon (7). Thus, the two cap sites correspond to a cytosine and adenine, respectively (Fig. 4). On the basis of the relative intensities of the observed bands we introduced a numbering scheme in which position +1 denotes the cytosine of a putative ME mRNA cap site, regardless of the consensus sequence reported for the start sites, where the adenine flanked by pyrimidines is most frequently the transcriptional start site in eukaryotic genes (24).

An analysis for local secondary structures of the sequence upstream from the putative mRNAs cap site revealed the potential of stem–loop structure formation immediately upstream from positions +1 to −163. The four such structures are shown in Fig. 4. The free energy for the base-pair structure for each configuration varies; the most stable one [ΔG = −31.9 kcal/mol (15); 1 cal = 4.184 J] was obtained for residues −1 to −39. Thus, it could be argued that the length of the primer-extended products indicates only the position of a strong stop site for reverse transcriptase in copying the mRNA rather than the 5′ end of the mRNA. However, this is unlikely, because the deletion mutant, pMEEPS28CAT, with a 5′ boundary at position −41 when tested in transient transfection assays for ability to express chloramphenicol acetyltransferase either in mouse hepatoma or Chinese hamster ovary cell lines allowed expression of 40 or 12% of maximal level, respectively (19). To determine whether additional start sites are present upstream of those detected by primer extension, we performed S1 nuclease experiments. As a probe we used a 5′-labeled 223-bp *Nae* I–*Ava* I fragment from genomic clone λ29 (Fig. 3*C*). The ^{32}P-labeled probe was

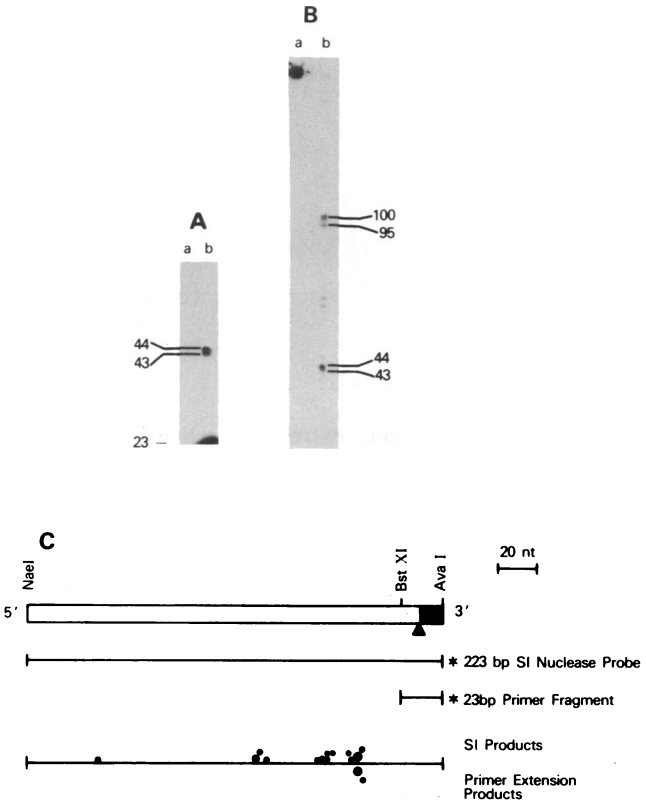

FIG. 3. Mapping of the 5′ end of the ME mRNA by primer-extension and S1 nuclease analyses. Poly(A)$^+$ RNA (25 μg) from thyroid hormone-treated rat liver was hybridized to ^{32}P-labeled *Bst*XI–*Ava* I fragment or *Nae* I–*Ava* I fragment (*C*) under appropriate conditions and reactions were incubated. The products were analyzed on sequencing gels followed by autoradiography. The *Bst*XI and *Nae* I sites are located upstream of the translation initiation codon, designated as a solid triangle, *Ava* I is present in the coding region, shown as a solid bar, and the asterisks indicate the 5′ end label of the noncoding strand (*C*). (*A*) For the primer-extension assay, 2 × 10^5 cpm of the *Bst*XI–*Ava* I fragment was incubated with yeast tRNA (lane a) and rat liver poly(A)$^+$ RNA (lane b) with 5 units of reverse transcriptase. Numbers indicate the sizes of the probe and extended products, determined from sequencing reaction of the cDNA fragment. (*B*) For S1 nuclease mapping analysis, 6.2 × 10^4 cpm of the *Nae* I–*Ava* I fragment was incubated with yeast tRNA (lane a) or rat liver poly(A)$^+$ RNA (lane b), then digested with 48 units of S1 nuclease. Numbers indicate the length of some of the S1-resistant fragments, using the labeled *Msp* I digest of pBR322 as size standards. (*C*) Summary of S1 and primer-extension products is shown, where the size of dots represent the relative intensities of the products.

hybridized to liver poly(A)$^+$ RNA from rats treated with thyroid hormone and incubated with S1 nuclease. Multiple protected fragments were detected (Fig. 3 *B* and *C*) with lengths corresponding to the 5′ untranslated regions of 30, 31, 46–53, 81, 85, 86, and 172 nt, respectively. The relative intensities at nt 30 and 31 (relative to the ATG) agreed between the primer extension and S1 nuclease protection experiments, suggesting these are the major cap sites. S1 nuclease resistant fragments identifying with the sizes of untranslated regions 81, 85, and 86 nt ending with the bases adenine, adenine, and cytosine, respectively, were less abundant. In addition, we observed several other bands in this assay that could represent additional minor transcript initiation sites. Thus, the 5′ end of the ME transcript exhibits a length heterogeneity as determined by S1 nuclease and primer-extension experiments.

Interestingly, there are no TATA or CCAAT box homologies in the vicinity of any transcriptional start sites (Fig. 4).

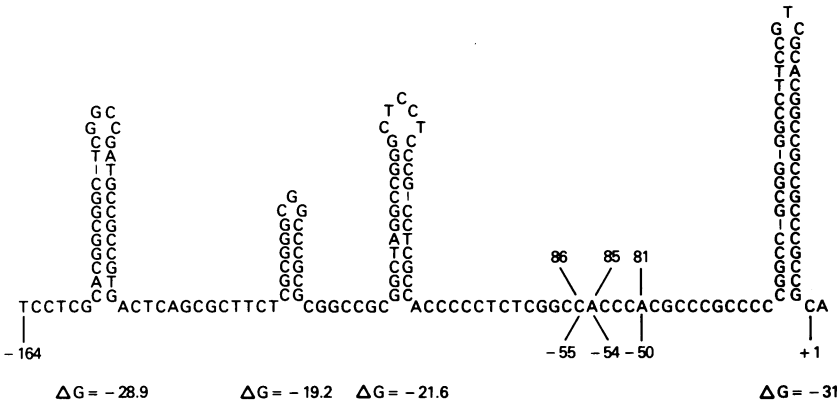

FIG. 4. Secondary structures of a 164-nt region in the 5′ end of ME gene. Four potential configurations, using a computer program, are shown for the sequences determined from the genomic DNA as indicted in Fig. 2. The free energy values in kcal/mol were calculated by the same computer program. The numbers below the sequence refer to the position of nucleotides relative to the strongest transcription start site (+1) as determined by primer-extension and S1 nuclease analysis. The numbers above the sequence designate the length of the untranslated regions relative to the adenine in the translation initiation codon, position 2.

This region is, however, extremely rich in guanine and cytosine (>83%). It has been recognized that genes lacking these sequences initiate transcription at multiple sites (for review, see ref. 25), probably due to an imprecision in the polymerase nucleotide selection mechanism. Many of these genes code for enzymes that are expressed at low levels in many tissues, as for ME, and are termed "housekeeping genes."

The 3′ End of the ME Gene. Rat ME mRNAs exist as two species of different sizes and their relative ratio varies in a tissue-specific manner (26). The nucleotide sequence for the shorter mRNA has been determined (7) and contains the polyadenylylation signal AATAAA found in most eukaryotic mRNAs (4). To identify the additional sequences contained in the larger ME mRNA we screened a rat liver λgt11 cDNA library with a fragment of genomic DNA located on the 3′ side of the end of the shorter mRNA. The clone obtained, λME36 (Fig. 2), overlapped with the sequence of the shorter mRNA and terminated at an internal *Eco*RI site. The sequence of the genomic clone, λ1 (Fig. 1), downstream of the coding region presented in exon 14, matches exactly with the sequence of λME36 and contains the second polyadenylylation signal ATTAAA, found in about 10% of eukaryotic mRNAs. These results indicate that the size differences of the two ME mRNA are due to heterogeneity at the 3′ untranslated region and originate from the use of alternative polyadenylylation signals. These results were confirmed by S1 nuclease mapping (data not shown). Thus, it appears that the mechanism by which these two ME mRNAs are synthesized in rat and mouse is similar (3).¶

¶An error exists in the nucleotide sequence of rat ME cDNA at nt 1704 (7); removal of an extra thymine residue at this position results in a protein containing 572 predicted amino acid residues, as has been reported for mouse ME (27).

We express our thanks and appreciation to Valerie Christy who assisted in some of the exon sequence analysis and to V. R. de Grange for editorial assistance and manuscript preparation.

1. Dozin, B., Rall, J. E. & Nikodem, V. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4705–4709.
2. Song, M.-K., Dozin, B., Grieco, D., Rall, J. E. & Nikodem, V. M. (1988) *J. Biol. Chem.*, in press.
3. Magnuson, M. A. & Nikodem, V. M. (1983) *J. Biol. Chem.* **258**, 12712–12717.
4. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
5. Graves, B. J., Johnson, P. F. & McKnight, S. L. (1986) *Cell* **44**, 565–576.
6. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
7. Magnuson, M. A., Morioka, H., Tecce, M. F. & Nikodem, V. M. (1986) *J. Biol. Chem.* **261**, 1183–1186.
8. Ullrich, A., Berman, C. H., Dull, T. J., Gray, M. & Lee, J. M. (1984) *EMBO J.* **3**, 361–364.
9. Sanger, F., Niklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
12. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217**, 316–324.
13. Sazer, S. & Schimke, R. T. (1986) *J. Biol. Chem.* **261**, 4685–4690.
14. Tinoco, I. J., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **248**, 40–41.
15. Blake, C. (1983) *Nature (London)* **306**, 535–537.
16. Bass, F., van Ommen, G.-J. B., Bikker, H., Arnberg, A. C. & de Vijlder, J. J. M. (1986) *Nucleic Acids Res.* **14**, 63.
17. Wyman, A. R., Wolfe, L. B. & Botstein, D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2880–2884.
18. Gitchier, J., Wood, W. I., Govalka, 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.
19. Morioka, H., Tennyson, G. E. & Nikodem, V. M. (1988) *Mol. Cell. Biol.* **8**, 3542–3545.
20. Towle, H. C., Mariash, C. N., Schwartz, H. L. & Oppenheimer, J. H. (1981) *Biochemistry* **20**, 3486–3492.
21. Lewin, B. (1983) *Genes* (Wiley, New York).
22. Henikoff, S., Keene, M. A., Fechtel, K. & Fristrom, J. W. (1986) *Cell* **44**, 33–42.
23. Beale, E., Chrapkiewicz, N., Scoble, H., Metz, R., Quick, D., Noble, R., Donelson, J., Biemann, K. & Granner, D. (1985) *J. Biol. Chem.* **260**, 10748–10760.
24. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
25. Dynan, W. S. (1986) *Trends Genet.* **2**, 196–197.
26. Dozin, B., Magnuson, M. A. & Nikodem, V. M. (1986) *Biochemistry* **24**, 5581–5586.
27. Bagchi, S., Wise, L. S., Brown, M. L., Bregman, D., Sul, H. S. & Rubin, C. S. (1987) *J. Biol. Chem.* **262**, 1558–1565.